Image features for pixel-wise detection of solar photovoltaic arrays in aerial imagery using a random forest classifier

Abstract—**Power generation from distributed solar photovoltaic (PV) arrays has grown rapidly in recent years. As a result, there is interest in collecting information about the quantity, power capacity, and energy generated by such arrays; and to do so over small geo-spatial regions (e.g., counties, cities, or even smaller regions). Unfortunately, existing sources of such information are dispersed, limited in geospatial resolution, and otherwise incomplete or publically unavailable. As result, we recently proposed a new approach for collecting such distributed PV information that relies on computer algorithms to automatically detect PV arrays in high resolution aerial imagery** [1]**. Here, we build on this work by investigating a new PV detection algorithm based on a Random Forest (RF) classifier, and we consider its detection performance using several different sets of image features. The proposed method is developed and tested using a very large collection of publicly available** [2] **aerial imagery, covering 112.5 km^2 of surface area, with 2,328 manually annotated PV array locations. The results indicate that a combination of local color and texture (using the popular texton feature) features yield the best detection performance.**

Keywords-component; convolutional neural networks, deep learning, detection, solar, energy, photovoltaic

I. INTRODUCTION

The quantity of solar photovoltaic (PV) arrays has grown rapidly in the US in recent years [3], [4], and a large proportion of this growth is due to small-scale, or distributed, PV arrays [5], [6]. As a result, there is growing interest among government agencies, utilities, and third party decision makers in having access to detailed information about distributed PV. Information such as the locations, power capacity, and the energy production of existing arrays is helpful for strategic energy-related decisions. Unfortunately, existing methods of obtaining this information, such as surveys and utility interconnection filings, are costly and time consuming to collect. They are also typically limited in spatial resolution to the state level, or larger [3],[6].

Based on these challenges, we recently proposed a new approach for collecting PV information [1]. This approach relies on using computer algorithms to automatically identify PV arrays in high resolution (≤ 0.3 meter) aerial imagery. Fig. 1a shows an example of 0.3 meter resolution aerial imagery in which the PV arrays have been annotated. This approach permits the collection of PV array information at a very high geo-spatial resolution. Also, because the approach is largely automatic, it is inexpensive to apply, and to do so repeatedly as new imagery is collected.

We build on previous work by investigating the use of a Random Forest classifier [8] (RF) for detecting PV arrays in aerial imagery. The RF is a popular supervised classification algorithm that has been used recently for a variety of problems [9]–[12], including the classification of aerial imagery [13]–[16]. Here the RF is used to classify each pixel in the imagery as corresponding to PV or not. Pixel-wise classification is useful for estimating the precise shape and size of PV arrays. These estimates can then be used to better estimate the capacity and power generation of the PV arrays.

In addition to considering alternative classifiers, another goal of this work is to investigate effective image features for the RF classifier. In binary classification tasks, such as the one considered here, the objects are typically represented by a feature vector. This is a vector of values that encodes information about each object. Image features, for example, often encode local color, texture, and shape information that might be indicative of which type of object is present. Here we conduct experiments with several successful image features, and compare their performance.

The performance of the RF on these proposed features is evaluated using a large dataset of aerial imagery encompassing $112.5\,\mathrm{km}^2$ of surface area, and 2,328 PV arrays. The true locations of PV arrays in the imagery have been manually annotated. Our data is a subset of a large collection of publicly available data, available here [2]. The experimental results demonstrate that the RF achieves excellent PV array detection accuracy, and that a combination of features encompassing color and grayscale texture perform best.

The remainder of this paper is organized as follows. Section II describes the aerial imagery dataset. Sections III and IV present the RF and image features, respectively. Section V describes the experimental design and the results; and Section VI presents our conclusions and ideas for future work.

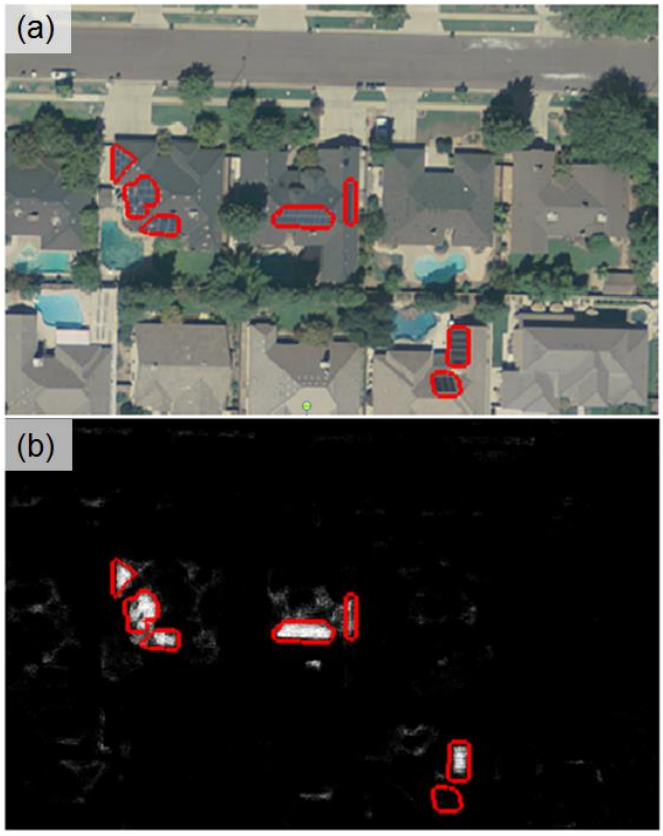

Fig. 1 (a) provides an example of the color orthoimagery used in this work along with several examples of solar PV annotations (red polygons). (b) is the pixel-wise output of the RF detector along with the same red polygons shown in (a). Bright locations indicate regions of high confidence, where PV arrays are likely to exist.

II. AERIAL IMAGERY DATASET

We conducted all experiments in this work using a dataset of color (RGB) aerial imagery, collected over the U.S. city of Fresno, California. All of the imagery was collected in the same month in 2013, using ortho-rectified aerial photography, with a spatial resolution of 0.3 meters per pixel. An example of the imagery is shown in Fig. 1a, where the solar PV locations are annotated in red. The locations of PV arrays in the aerial imagery were manually annotated with polygons. The full dataset used in this work encompasses 112.5 km^2 of surface area, and 2,328 PV array annotations. The dataset we used in the experiments is a random subset of the larger set of imagery available in [2], where it can be downloaded freely.

In order to avoid a positive bias from overfitting in the performance evaluation of the proposed PV detection algorithms, we split our experimental dataset into two disjoint sets: Fresno Training and Fresno Testing. This validation set approach is common for evaluating the performance of supervised machine learning algorithms, such as those considered here. A summary of the imagery in each dataset is presented below in Table 1.

TABLE 1
SUMMARY OF FRESNO COLOR ORTHOIMAGERY DATASET

Designation	Area of Imagery	Number of PV Annotations
Fresno Training	90.0 km^2	1780
Fresno Testing	22.5 km^2	548

III. THE RANDOM FOREST PV ARRAY DETECTOR

Random Forests (RF) [8] are a supervised statistical classification method that has been successfully applied to a variety of problems [9]–[15]. In this work the RF performs detection in a pixel-wise basis, by assigning a "confidence" to each pixel indicating how likely it is that the pixel corresponds to a solar PV array. The result of this processing is a confidence map, indicating where PV arrays are likely to exist. An example image and its corresponding RF confidence map are shown in Fig. 1.

Because the RF is a supervised classification model, it automatically infers most of its parameters using training data. The training data consists of pairs of feature vectors, and their corresponding labels (PV or non-PV). The RF actually consists of an ensemble of T simpler supervised classifiers called decision trees [17]. Employing more trees improves performance, but increases computational costs during training and testing. In this work $T = 30$ was found to achieve a good balance between these factors, based on the RF performance on the Fresno Training data.

During training, each tree is grown independently of the other trees, in a top-down manner, using a random bootstrap sample of pixels from the training data. Each tree consists of a series of decision nodes that are learned such that they best separate the training data according to some performance measurement. In this work we use the Gini index, which is commonly used for the RF. Each node of each tree considers only a random subset of the input features (of size m) when inferring how to split the data. The parameter m is often cited as the only major adjustable parameter of the RF. A conventional setting of m that usually works well is $m = \sqrt{M}$, where M is the number of feature dimensions [14].

Each tree terminates (at the bottom of the tree) in a leaf node. At the leaf node a probability is assigned to the input feature vector indicating the probability that it corresponds to a PV array.

IV. IMAGE FEATURES FOR THE CLASSIFIER

In this work, we investigate several sets of features as input for the RF pixel-wise classifier, which are described next.

A. *Raw pixels*

The raw pixel features simply consist of the intensities of the pixels surrounding the pixel we wish to classify. When proposing new features for a machine learning problem, it is useful to prove that the proposed features improve over simpler approaches that require less processing. Raw pixels are employed in this work to provide a baseline (or benchmark) performance for the other more sophisticated

features that are considered. Raw pixels are perhaps the simplest possible features, because they require no additional processing of the imagery.

Determining the window size around the pixel under consideration is another important algorithm design issue. Using a bigger window captures more local image information, which can be useful for classification. However, increasing the window size also increases the dimensionality of the feature vector. This can lead to the classifier overfitting the data, and therefore poor performance. A 7x7 window was chosen to balance these two considerations. In this work raw pixels were extracted in a 7x7 window around each pixel location. This leads to a total of 147 total features (49 features for each channel of the RGB imagery).

B. *Local Color Statistics (LCS)*

The local color statistics (LCS) feature has recently been used [18], [19] as a computationally inexpensive method of encoding local color information, as well as texture. The LCS feature vector consists of the means, μ, and variances, σ^2, of pixel intensities, computed in windows surrounding the pixel to be classified.

The window size of the LCS feature was chosen to be 3x3, because this size roughly corresponds to the size of the smallest PV array in the dataset. A larger window size would risk mixing PV pixels with background pixels, and thereby obscuring the information useful for identifying individual PV panels.

The full LCS feature vector consists of the statistics from many windows surrounding the pixel we wish to classify, denoted here as p_0. The windows are organized into two rings surrounding p_0, and each ring consists of 9 windows. Each window can be characterized by its vertical offset, y, and horizontal offset, x, from p_0. This is illustrated in Fig. 2.

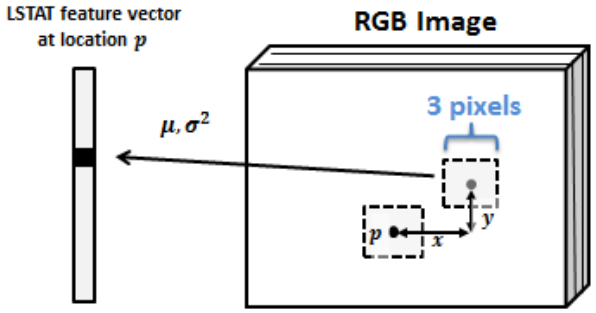

Fig. 2 (a) provides an example of the color ortho-imagery used in this work, along with several examples of solar PV annotations (red polygons). (b) is the pixel-wise output of the RF detector along with the same red polygons as in 1a. Bright locations indicate regions of high confidence, where PV arrays are likely to exist.

A set of 9 windows (one ring) is denoted here by S_r, where the subscript r parameterizes the locations of the windows in the ring. The locations of the windows in S_r are then given by

$$S_r = \{(x, y): x \in \{0, -r, r\}, y \in \{0, -r, r\}\}. \quad (1)$$

There are 9 windows in each ring, and each window yields 6 features (a μ and σ^2 for each color channel), resulting in 54 total features per ring. In this work, we extracted features in two rings, given by S_2 and S_4. Note that S_2 and S_4 share a window location at $(x, y) = (0,0)$. One of these duplicates is removed, leaving a total of 54+54-6=102 total features.

C. *Textons*

Texton features are a popular class of features that are designed to capture image texture information [20]. Textons have been used successfully for recognition in images [20]–[23], as well as aerial imagery [24], [25]. In contrast to the other features investigated in this work, textons require a training step to learn a database of textures and shapes, called a dictionary. The entries in the dictionary are referred to as textons. The texton training procedure is illustrated and described in Fig. 3.

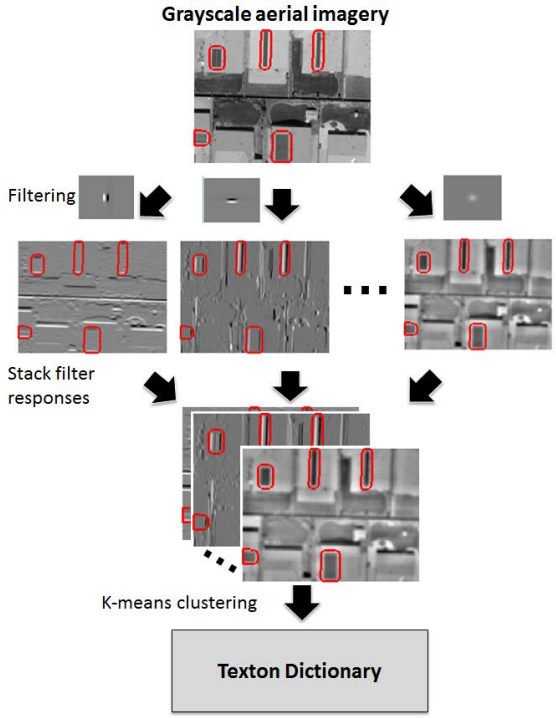

Fig. 3 An illustration of the texton feature training procedure. Aerial imagery is filtered with a bank of filters. In this work the Leung-Malik (LM) filter bank is used [30], with 6 orientations, and 2 scales (36 total filters). The filtered images are stacked together, where each pixel is represented by a vector of 36 values, corresponding to its filter responses. The vector of filter responses for each pixel is then used as input to a K-means clustering algorithm, which learns K representative, or common, filter response vectors, which are called textons in this context. In this work K was set to 30. The textons are learned on gray-scale imagery rather than the original color imagery.

In this work a million pixels from the Fresno Training dataset are used for learning the texton dictionary. All of the PV array pixels are used (roughly 500,000) and the remaining pixels are sampled randomly from non-PV locations. The texton features in this work are trained on and extracted from, aerial imagery that has been converted from RGB color to grayscale. This is intended to force the texton features to encode only texture and shape information, as opposed to

color information. This makes the textons more complementary to the LCS features. In Section V, we show that the best performance is achieved by combining texton and LCS features.

The extraction of texton features for classification is illustrated in Fig. 4. Essentially, the feature vector for a given pixel consists of a histogram of the textons that were present in the imagery surrounding that pixel. This is a compact characterization of the local textures and shapes. In this work, the histograms were created using 9x9 windows centered on the pixel location that is being represented. This window size was chosen to be large enough to capture several local PV array textures, but small enough that these textures are not usually diluted by the presence of non-PV textures from the background.

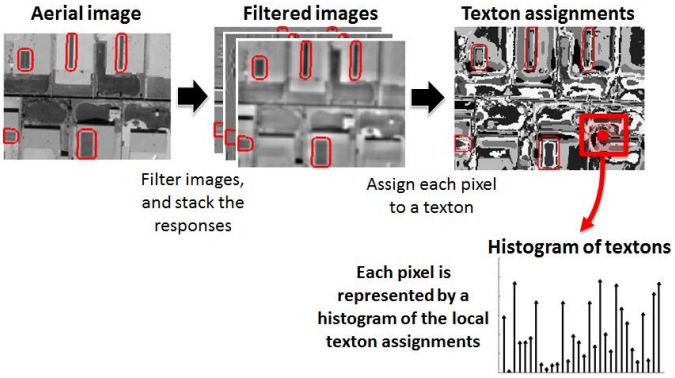

Fig. 4. Illustration of the process of extracting texton features for a patch of aerial imagery. Each image is filtered with the LM filter bank. Each pixel is assigned to a texton based on its responses to the LM filters. This results in an image where each pixel is represented by its texton assignment. There are thirty textons, so each pixel receives a value between one and thirty. The feature vector for a given pixel consists of a histogram of the texton assignments in a centered 9x9 window.

V. Experiments and Results

This section presents the design of the experiments in this work, followed by a discussion of the results.

A. Experimental Design

A total of three million pixels from the Fresno Training imagery were used to train the RF classifier. All of the available PV pixels were used (roughly 500,000), and the remaining training pixels were sampled randomly from the non-PV imagery. The primary role of the Fresno training dataset was to train the RF classifier and texton dictionary. The Fresno Testing dataset was used to obtain an unbiased performance estimate for the detector. This is a common cross-validation approach for supervised machine learning algorithms [26].

The metric used to evaluate the performance of the algorithm is the precision recall (PR) curve. The PR curve is a popular performance metric for object detection in aerial imagery [16], [27]–[29], and therefore it is adopted here. PR curves measure the performance tradeoff between making correct detections and false detections, as the sensitivity of a detector is varied. The x-axis of a PR curve is the recall, R, which is the proportion of all true target objects (i.e., PV arrays) in the data that were returned by the algorithm as detections. The y-axis is the precision, P, which is the proportion of all detected objects (i.e., both true and false) which are true targets. An effective detector will have a PR curve that tends towards the top right corner of the PR curve space, where both recall and precision are maximized.

B. Results

The results of applying the trained RF to the Fresno Testing dataset are shown in Fig. 5. First, let us consider the two features that use color: raw pixels and LCS. The results indicate that the LCS features lead to substantially better performance compared to the performance achieved with raw pixels. As discussed in Section III, this result is expected, because raw pixels represent a simple baseline feature. Nonetheless, it confirms that the extra processing required for the LCS features results in substantially improved performance.

The results show that the single worst performing feature is the (grayscale) textons. This is not surprising because color is an important cue for the presence of a PV array, and the textons do not use it. The textons used here operate on grayscale imagery, and therefore are forced to rely only on texture. This was a deliberate design choice in order to make the textons more complementary to the LCS features. The results indicate that the best performing feature set consists of both the LCS and texton features. Although the textons perform poorly when used alone, they yield improvements for the LCS feature due to their complementarity (i.e., they do not encode redundant information).

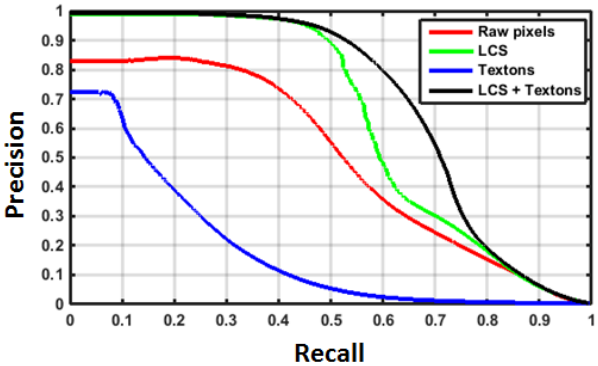

Fig. 5. Precision-recall curves for the RF classifier for the different feature sets investigated in this work.

VI. Conclusions

In this work we investigated the utilization of a supervised RF classifier for performing pixel-wise detection of PV arrays in aerial imagery. The performance of the RF was evaluated with several different sets of features: raw pixels, LCS, and grayscale textons. The results indicate that the LCS features outperform the other features individually, but that the best overall performance is achieved by combining the texton and LCS features together. This suggests that, consistent with intuition, texture and color information are both important cues for identifying PV array pixels.

The improved performance obtained by combining the textons and LCS features is achieved at the cost of some additional computation during feature extraction (e.g., to apply filters, or create histograms). Future work may design texture and color features that have lower computational costs. Additionally, encoding more contextual information into features could enhance classifier efficacy, such as features based on the building rooftop pixels that surround most solar panels in urban or suburban settings.

REFERENCES

[1] J. M. Malof, R. Hou, L. M. Collins, K. Bradbury, and R. Newell, "Automatic solar photovoltaic panel detection in satellite imagery," in *International Conference on Renewable Energy Research and Applications (ICRERA)*, 2015, pp. 1428–1431.

[2] K. Bradbury, R. Saboo, J. Malof, T. Johnson, A. Devarajan, W. Zhang, L. Collins, and R. Newell, "Distributed Solar Photovoltaic Array Location and Extent Data Set for Remote Sensing Object Identification," *figshare*, 2016. [Online]. Available: https://dx.doi.org/10.6084/m9.figshare.3385780.v1. [Accessed: 01-Jun-2016].

[3] M. J. E. Alam, K. M. Muttaqi, and D. Sutanto, "An approach for online assessment of rooftop solar PV impacts on low-voltage distribution networks," *IEEE Trans. Sustain. Energy*, vol. 5, no. 2, pp. 663–672, 2014.

[4] A. Chersin, W. Ongsakul, J. Mitra, and S. Member, "Improving of Uncertain Power Generation of Rooftop Solar PV Using Battery Storage," in *International Conference and Utility Exhibition on Green Energy for Sustainable Development*, 2014, no. March, pp. 1–4.

[5] "Electric Power Monthly," US Energy Information Administration, 2016.

[6] "Net Generation from Renewable Sources: Total (All Sectors), 2006-February 2016," US Energy Information Administration, 2016.

[7] L. Sherwood, "U.S. Solar Market Trends 2013," 2013.

[8] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001.

[9] V. Lepetit and P. Fua, "Keypoint recognition using randomized trees," *IEEE Trans. Pattern Anal. Mach. Intell.*, 2006.

[10] V. Lempitsky, M. Verhoek, J. A. Noble, and A. Blake, "Random forest classification for automatic delineation of myocardium in real-time 3D echocardiography," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5528, pp. 447–456, 2009.

[11] G. Fanelli, J. Gall, and L. Van Gool, "Real time head pose estimation with random regression forests," *IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 617–624, 2011.

[12] J. Shotton, a W. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, a Kipman, and a Blake, "Real-time human pose recognition in parts from single depth images," *Cvpr*, pp. 1297–1304, 2011.

[13] P. Tokarczyk, J. Montoya, and K. Schindler, "an Evaluation of Feature Learning Methods for High Resolution Image Classification," *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, vol. I–3, pp. 389–394, 2012.

[14] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, "Random forests for land cover classification," *Pattern Recognit. Lett.*, vol. 27, no. 4, pp. 294–300, 2006.

[15] M. Pal, "Random forest classifier for remote sensing classification," *Int. J. Remote Sens.*, vol. 26, no. 1, pp. 217–222, 2005.

[16] a M. Cheriyadat, "Unsupervised Feature Learning for Aerial Scene Classification," *Geosci. Remote Sensing, IEEE Trans.*, vol. 52, no. 1, pp. 439–451, 2014.

[17] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and regression trees*. CRC press, 1984.

[18] S. Clinchant, G. Csurka, F. Perronnin, J. Renders, C. Europe, and D. Maupertuis, "XRCE ' s participation to ImagEval."

[19] J. Sánchez, F. Perronnin, T. Mensink, and J. Verbeek, "Image Classification with the Fisher Vector: Theory and Practice," *Int. J. Comput. Vis.*, vol. 105, no. 3, pp. 222–245, 2013.

[20] M. Varma and A. Zisserman, "A statistical approach to texture classification from single images," *Int. J. Comput. Vis.*, vol. 62, no. 1–2, pp. 61–81, 2005.

[21] a Bosch, a Zisserman, and X. Munoz, "Scene classification using a hybrid generative/ discriminative approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 4, pp. 712–727, 2008.

[22] T. Leung and J. Malik, "Representing and Recognizing the Visual Appearence of Materials using Three-dimensional Textons," *Int. J. Comput. Vis.*, vol. 43, no. 1, pp. 29–44, 1999.

[23] J. Kaufhold, "Recognition and Segmentation of Scene Contentusing Region-Based Classification," no. middle L, 2006.

[24] S. Ø. Larsen, A. Salberg, and R. Solberg, "Automatic avalanche mapping using texture classification of optical satellite imagery," 2013.

[25] D. R. Patlolla, S. Voisin, H. Sridharan, and a M. Cheriyadat, "GPU Accelerated Textons and Dense SIFT Features for Human Settlement Detection from High - Resolution Satellite Imagery," pp. 1–7, 2012.

[26] C. M. Bishop, *Pattern recognition and machine learning*, vol. 1. springer New York, 2006.

[27] D. Chaudhuri, N. K. K. Kushwaha, A. Samal, and R. C. C. Agarwal, "Automatic Building Detection From High-Resolution Satellite Images Based on Morphology and Internal Gray Variance," *Sel. Top. Appl. Earth Obs. Remote Sensing, IEEE J.*, vol. PP, no. 99, pp. 1–13, 2015.

[28] A. Manno-Kovacs and A. O. Ok, "Building Detection From Monocular VHR Images by Integrated Urban Area Knowledge," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 10, pp. 2140–2144, 2015.

[29] V. Mnih and G. E. Hinton, "Learning to detect roads in high-resolution aerial images," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6316 LNCS, no. PART 6, pp. 210–223, 2010.

[30] J. Malik, S. Belongie, T. K. Leung, and J. Shi, "Contour and Texture Analysis for Image Segmentation," *Int. J. Comput. Vis.*, vol. 43, no. 1, pp. 7–27, 2001.