

MAN-MADE OBJECT EXTRACTION FROM REMOTE SENSING IMAGERY BY GRAPH-BASED MANIFOLD RANKING

Y. He^{1,*}, X. Wang¹, X. Y. Hu², S. H. Liu¹

¹ Satellite Surveying and Mapping Application Center, National Administration of Surveying, Mapping and Geo-information, Beijing, P.R. China - (hey, wangx, liush)@sasmac.cn

² Collaborative Innovation Center of Geospatial Technology, Wuhan University, Wuhan, P.R. China - huxy@whu.edu.cn

Commission III, WG III/1

KEY WORDS: Man-made Object Extraction, Remote Sensing Image, Priori Area Extraction, Graph Model, Manifold Ranking

ABSTRACT:

The automatic extraction of man-made objects from remote sensing imagery is useful in many applications. This paper proposes an algorithm for extracting man-made objects automatically by integrating a graph model with the manifold ranking algorithm. Initially, we estimate a priori value of the man-made objects with the use of symmetric and contrast features. The graph model is established to represent the spatial relationships among pre-segmented superpixels, which are used as the graph nodes. Multiple characteristics, namely colour, texture and main direction, are used to compute the weights of the adjacent nodes. Manifold ranking effectively explores the relationships among all the nodes in the feature space as well as initial query assignment; thus, it is applied to generate a ranking map, which indicates the scores of the man-made objects. The man-made objects are then segmented on the basis of the ranking map. Two typical segmentation algorithms are compared with the proposed algorithm. Experimental results show that the proposed algorithm can extract man-made objects with high recognition rate and low omission rate.

1. INTRODUCTION

The automatic extraction of man-made objects, such as buildings, roads and bridges, from remote sensing images is one of the fundamental but challenging tasks in remote sensing and computer vision. It has been applied in many applications, including urban planning (Hinz and Baumgartner, 2003), disaster evaluation (Turker and Sumer, 2008), change detection (Ji and Yuan, 2007) and map updating (Florczyk et al., 2016). In recent decades, remote sensing images have been widely used in man-made object extraction. However, it is still difficult to recognize targets from the images with complex structures completely.

State-of-the-art algorithms for man-made object extraction can be divided into two categories: feature-based and model-based algorithms. Many existing studies have utilized feature-based algorithms, which focus on the low-level visual features of images. Levitt and Aghdasi (1997) used texture features and introduced the homogeneous operator. However, this operator is unsuitable for regions with heterogeneous man-made objects. In the method developed by Wang and Yang (2010), extraction is accomplished by analysing morphological features. On the basis of human visual attention mechanism, Cai et al. (2011) utilized texture and geometric structure features to extract targets. In the process of extraction, feature-based algorithms directly focus on the image feature to be extracted instead of the category of the target. Consequently, these methods have extensive applicability to both single-object and complex scenes.

Although feature-based algorithms have gained considerable success, they are still far from satisfactory because of low-level

knowledge, which results in the process of segmentation and target extraction blindly. On the basis of knowledge rules, model-based algorithms analyse the characteristics of the target in the image to facilitate the process of segmentation. Li et al. (2000) utilized a 2D hidden Markov model and estimated the values of the relevant parameters by an EM algorithm. Reno and Booth (1999) extracted man-made objects by using a viewer-centred reference model with a deformable template on the basis of combining model and image information. Model-based algorithms present high pertinence and efficiency, thereby avoiding extracting numerous unnecessary features in the image. However, once the target has changed, the corresponding knowledge rules have to be changed; thus, model-based algorithms are less versatile. In addition, targets can be extracted by deep learning algorithms (Makantasis et al., 2015), which usually requires labeled samples for training the deep neural network.

The algorithm proposed in this paper is a feature-based algorithm. We have observed that man-made objects often present symmetric appearances as well as high contrast with adjacent land cover types. Thus, we utilize these cues in this work to extract a priori value of man-made objects from remote sensing images, and optimize it on the basis of multiple features to obtain the final result. We model man-made object detection as a manifold ranking problem with a graph, in which each node is a superpixel. The proposed approach consists of two stages. Figure 1 shows the main stages of the extraction algorithm for man-made objects. In the first stage, a contrast prior is extracted using the entropy of the histogram and genetic quantum algorithm. Meanwhile, the images are divided into superpixels, and whether a superpixel is a symmetric part is determined by

* Corresponding author. Tel.: +86-10-68412381; fax: +86-10-68412678.
E-mail address: hey@sasmac.cn (Y. He).

machine learning to obtain a symmetry prior. These two prior images are then integrated to generate a man-made object priori map, from which initial seeds are generated for the manifold ranking algorithm. In the second stage, the seeds from the first stage are taken as the man-made object queries, and the graph-based manifold ranking is used to optimize the prior priori map. The graph is constructed using superpixels, and the colour, texture and main direction features are modelled in the graph. The probability of each node being a part of a man-made object is computed on the basis of its relevance to the a priori queries. The extensive experiments conducted validate the high efficiency of the proposed approach in man-made object extraction.

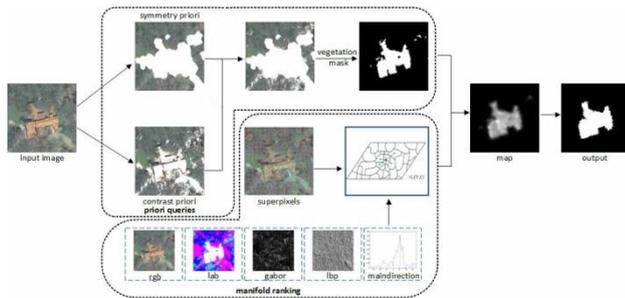


Figure 1. Diagram of the proposed model

2. GRAPH-BASED MANIFOLD RANKING

The graph-based manifold ranking model is a process of spreading labels from the given queries to the remaining nodes. Essentially, it is a graph-based semi-supervised method. Thus, the ranking algorithm aims to construct a weighted graph and identify the relevance between unlabelled nodes and queries (Yang et al., 2013).

2.1 Manifold ranking

Manifold ranking, which was first proposed by Zhou et al. (2003) for graph labelling, is based on the intrinsic manifold structure of data. Given the dataset $V = \{v_1, v_2, \dots, v_p, v_{p+1}, \dots, v_n\} \in R^{m \times n}$, the first p points are the labelled queries, whereas the rest of the points need to be ranked. Let $s: V \rightarrow R^n$ be the ranking function that specifies the ranking score of each point. The vectors $s = [s_1, s_2, \dots, s_n]^T$ and $y = [y_1, y_2, \dots, y_n]^T$ are defined as the ranked and labelled vectors respectively. If v_i is a query, then $y_i = 1$; otherwise, $y_i = 0$. Next, a graph $G = (V, E, W)$, where the edge E is weighted by the affinity matrix $W = w_{ij}_{n \times n}$, is defined. The degree matrix $D = \text{diag}\{d_{11}, d_{22}, \dots, d_{nn}\}$, where $d_{ii} = \sum_{j=1}^n w_{ij}$, can be obtained on the basis of the affinity matrix. The ranking model can be represented by the following cost function:

$$O(s) = \sum_{i,j=1}^n w_{ij} \left\| \frac{1}{\sqrt{d_{ii}}} s_i - \frac{1}{\sqrt{d_{jj}}} s_j \right\|^2 + \mu \sum_{i=1}^n \|s_i - y_i\|^2 \quad (1)$$

where μ is the parameter that controls the balance between the smoothness and the fitting constraints. That is, nearby points are likely to have the same label, but the points should not deviate considerably from the original label. The final ranking score s^* can be obtained by solving the following optimization problem:

$$s^* = \arg \min_s O(s) \quad (2)$$

To get the minimum value, the derivative of the function can be set to zero. After deduction, the ranking function can be written as

$$s^* = Ay \quad (3)$$

where

$$A = (D - \frac{1}{1+\mu} W)^{-1} \quad (4)$$

2.2 Graph construction

A superpixel contains extensive feature information, and it can reduce computation; thus, the image is initially over-segmented by the SLIC algorithm (Achanta et al., 2010). Consequently, a set of superpixels, which are the nodes in the graph, are generated. Multiple features are utilized to describe the region/superpixel i completely:

$$v_i = [L_i, A_i, B_i, G_i, P_i, D_i] \quad (5)$$

where (L_i, A_i, B_i) is the average value of superpixel colour in the LAB colour space, G_i and P_i are the average texture information (Gabor and LBP operators are used in the experiment) and D_i is the main direction of the gradient in a region.

As mentioned in Section 2.1, the graph $G = (V, E, W)$, where V is the set of nodes and E is a set of undirected edges weighted by the affinity matrix $W = w_{ij}_{n \times n}$, is constructed. The weight between the two adjacent nodes in the graph is defined as

$$w_{ij} = \exp\left(\frac{-\|v_i - v_j\|^2}{2\sigma^2}\right) \quad (6)$$

In this work, σ^2 is a constant that controls the strength of the value. The weights are computed using a variety of features as it has been shown to be effective in man-made object detection.

3. MAN-MADE OBJECT EXTRACTION

The proposed two-stage scheme is based on the graph-based manifold ranking model. As mentioned in Section 1, a priori map of the man-made object should be obtained from which object queries for optimization are generated.

3.1 Extraction of the priori map

The geometric structures and appearances of man-made objects are obvious; thus, the symmetry and contrast features are used to extract the object region automatically.

Traditional symmetry detection algorithms use maximal discs (Blum, 1967) to determine the locus of a symmetric target. In the current study, superpixels at multiple scales are adopted as deformable maximal disc hypotheses (Levinshtein et al., 2013). A superpixel that represents a good maximal disc hypothesis should conform to two key forms of perceptual grouping. The first is homogeneity; a superpixel region should present continuity in its appearance. The second is symmetry; the maximal disc bitangency can be replaced by two opposing parts of the boundary of a superpixel. If the superpixels are segmented too finely or too coarsely, then the opposing boundary cannot successfully capture a symmetric part. Thus,

superpixels should be generated at different scales (figure 2(b)) by using the normalized cuts algorithm (Shi and Malik, 2000) to ensure that a superpixel represents a good maximal disc hypothesis. Consequently, an affinity between two adjacent superpixels at a given scale is established. The affinity matrix $A(i, j)$ has two elements, namely A_{shape} and $A_{appearance}$. The former determines whether the distribution of the adjacent edges of the superpixels fits the actual boundary via SVM classifier, and the latter verifies the homogeneity of the intervening region. Finally, the shape and appearance affinities are combined using a logistic regressor, and the edge weight $w(i, j) = \frac{1}{A(i, j)}$ is acquired. Agglomerative clustering (Felzenszwalb and Huttenlocher, 2000) based on the resulting graph is utilized, and we obtain the symmetric parts (figure 2(c)) in the image.

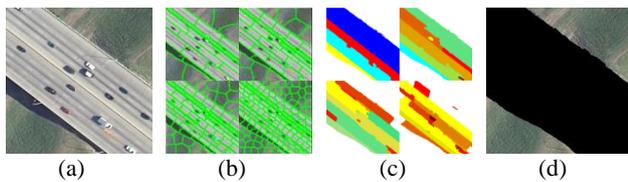


Figure 2. Steps of multiscale symmetric part detection: (a) original image, (b) multiscale superpixel segmentation, (c) symmetric part extraction at all scales and (d) generated symmetry prior.

Concerning the contrast feature of man-made object, target extraction is considered as a foreground segmentation problem, where the man-made object is regarded as the foreground region. The method proposed by Kapur et al. (1980), referred to as the KSW entropy method in the current paper, has achieved excellent results in solving image segmentation problems. However, it presents high computational overhead during for the determination an adaptive threshold. Therefore, the KSW entropy method based on quantum genetic algorithm (Gou, 2008) is applied in the current study to enhance computational efficiency and reduce the search time for the optimal threshold, with the goal of rapid extraction. In the KSW entropy method, the images are processed in greyscale, and the information distributed in the foreground and background of the experimental image is enabled to reach the maximum by using the entropy measurement of the grey histogram to determine the segmentation threshold automatically. As an optimization algorithm, the quantum genetic algorithm introduces the quantum coding form into chromosome encoding. By utilizing the inherent superposition, coherence and other characteristics of the quantum state, quantum genetic algorithm can represent multiple possible linear superposition states. Furthermore, quantum genetic algorithm exhibits rapid convergence and powerful global search capability in solving optimization problems because of quantum parallelism. Therefore, quantum genetic algorithm is combined with the KSW entropy method in the present study to enable the segmentation algorithm to global search, thus to find out the optimal segmentation threshold soon. After computing a priori value of symmetry and contrast, we obtain the final priori of man-made object for optimization.

We note that only depending on the contrast priori, some shadows and low-contrast areas may not be accurately detected as shown in figure 3(b). The multiscale symmetric part extraction, which focuses on the entire object, can reduce the effects of imprecise contrast queries (figure 3(c)). Therefore, the use of a fusion prior can facilitate the integral extraction of objects. However, some natural cover types (mainly vegetation)

are inevitably divided into symmetric parts. To decrease misjudgement rate, a vegetation mask by correlation between bands is utilized, and the final fusion prior is obtained (figure 3(d)).

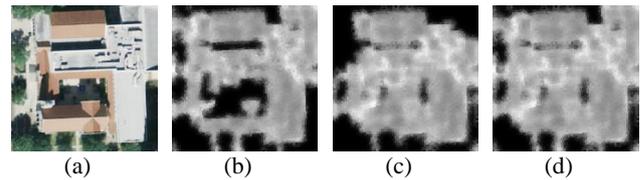


Figure 3. Man-made object extraction results using different queries: (a) original image, (b) result after using contrast prior as queries, (c) result of using symmetry prior as queries and (d) result of using fusion prior as queries.

3.2 Ranking with priori queries

Contrast and symmetry features, which facilitate the selection of the nodes of the priors as queries, are used to obtain the prior of the man-made object. The indicator vector \mathbf{y} can be formed after the queries are obtained. When the superpixel i in the priori image is a part of the man-made object, then $y_i = 1$; otherwise, $y_i = 0$.

A graph is constructed by connecting the adjacent superpixels generated by the SLIC algorithm. Since each node's property and appearance have a close correlation with the neighbouring nodes, superpixels are utilized as the nodes of the graph not only to reduce computation but also to consider the spatial correlation. Multiple features, namely the colour, texture and the main direction of the region, are integrated to describe a superpixel adequately. Using these features, the affinity matrix \mathbf{W} , which reflects the similarity between adjacent nodes, can be derived using equation (6). In particular, most of the elements of the matrix \mathbf{W} are zero because of the sparsely connected graph. On the basis of the indicator vector \mathbf{y} and affinity matrix \mathbf{W} , the final ranking map is derived by

$$V(i) = \bar{\mathbf{y}}^*(i) \quad i = 1, 2, 3, \dots \quad (7)$$

where i denotes the nodes on the graph and $V(i)$ represents the normalized ranking score. Finally, the extraction result of the man-made objects is acquired using a fixed threshold.

4. MAN-MADE OBJECT EXTRACTION

The extraction algorithm of man-made objects is tested using 20 remote sensing images. Our approach is evaluated and compared with the MRF model (Zong et al., 2015) and active contour approach (Wang et al., 2009). Our approach involves two parameters, namely μ , which controls the balance between the smoothness and fitting constraints in the optimization function, and the constant σ^2 for the affinity matrix \mathbf{W} . After conducting extensive experiments, we set $\mu = 0.7$ and $\sigma^2 = 0.5$ for the optimal results.

Figure 4 shows the detection results for an aerial image with 0.1 m resolution. Figures 4(b) and (c) show the a priori extraction result of the man-made object by using the symmetry and contrast features respectively. These results are integrated into a final priori result, which is shown in figure 4(d). After manifold ranking, the extraction results are obtained and shown in figures 4(e) and (f). The man-made objects detected by the MRF model and active contour approach are shown in figures 4(g) and (h)

for comparison. Clearly, the structures of the man-made objects in figure 4(a) are complex, and the shadows increase the difficulty of extraction. Nevertheless, our proposed approach can extract the areas more accurately than the two other approaches can, avoiding the effect of the shadows and texture noise. The results of the two existing approaches (figures 4(g) and (h)) present several insignificant and complex structure areas merged with the background.

Figure 5 shows the extraction results for an example (0.1 m resolution) from the remote sensing image dataset (UCMerced LandUse Dataset). The results of the proposed approach and the two other compared algorithms are roughly similar. However, some differences can still be observed in the regions enclosed in yellow frames. The highlighted regions are covered with buildings. In these regions, the contrast cues are similar to the background. Consequently, the results of the MRF and active contour approaches in figures 5(g) and (h) exhibit high fall-out ratios. By contrast, the result of our approach (figure 5(f)) shows that nearly all the man-made areas are properly detected. We utilize several characteristics and adjacency relation to reduce the rate of fault detection.

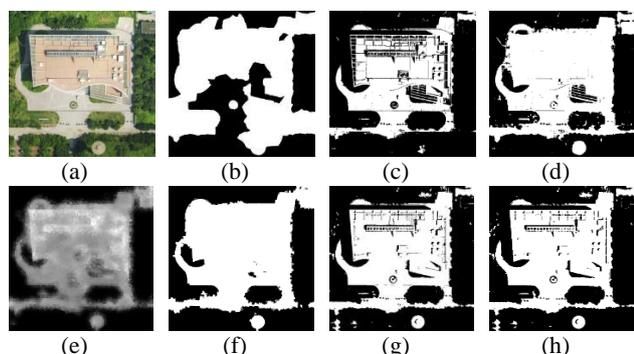


Figure 4. Man-made object extraction results for an aerial image: (a) original image, (b, c) a priori extraction results with symmetry and contrast features respectively, (d) a priori extraction result after fusion, (e, f) results of our method and binary image, (g) result obtained using the MRF algorithm by Zong et al. (2015) method, (h) result obtained using the active contour approach by Wang et al. (2009).

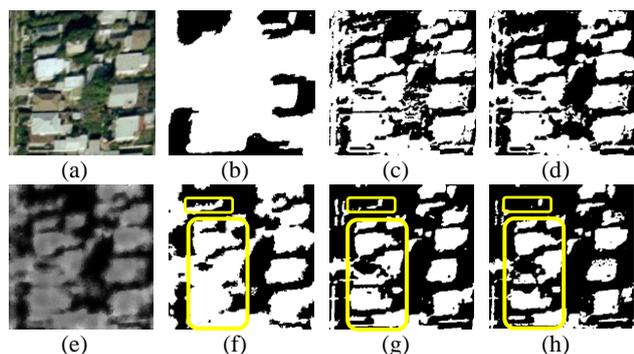


Figure 5. Man-made object extraction results for an aerial image: (a) original image, (b, c) a priori extraction results with symmetry and contrast features respectively, (d) a priori extraction result after fusion, (e, f) results of our method and binary image, (g) result obtained using the MRF algorithm by Zong et al. (2015) method, (h) result obtained using the active contour approach by Wang et al. (2009).

Figure 6 presents the results for RGB images of different scenes. These images are taken from different regions; thus, the

appearances of the man-made objects are diverse which obviously increase the difficulty of detection. Furthermore, the object types in these images include freeways, intersections, mobile home parks, tennis courts and buildings. Nevertheless, the results demonstrate that our approach performs fairly well for complex images.

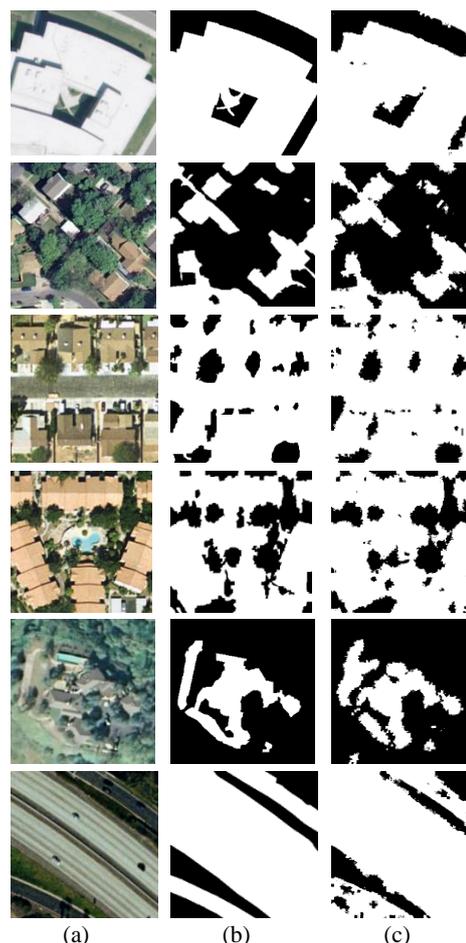


Figure 6. Man-made object extraction results for UCMerced LandUse dataset: (a) original images, (b) ground-truth images and (c) results of our method.

Following the method used by Achanta et al. (2009), we quantitatively evaluate the performance of our method in terms of precision, recall and F-measure. The precision value indicates the ratio of correctly assigned pixels that belong to man-made objects to all the pixels in the extraction area, and recall value is the proportion of the detected target pixels with respect to ground-truth number. After the extraction result is binarized using thresholds ranging from 0 to 255, the precision–recall curve can be acquired. Figure 7(a) shows the ranking results obtained using different features. Figure 7(b) presents the effects of the contrast and symmetry priors and indicates that using the fusion prior as queries outperforms using a single prior. F-measure is used to evaluate the performance of our algorithm and those of the two other approaches further. F-measure is computed using the weighted harmonic of precision and recall.

$$F_{\beta} = \frac{(1+\beta)\text{Precision} \times \text{Recall}}{\beta\text{Precision} + \text{Recall}} \quad (8)$$

where $\beta = 0.3$ according to Yang et al. (2013). Figure 7(c) demonstrates the precision, recall and F-measure values of the

three compared methods using the 20 test images. The chart shows that their precision values are approximate; however, our approach achieves the highest recall and F-measure values. Overall, the proposed algorithm is superior to the two other existing methods.

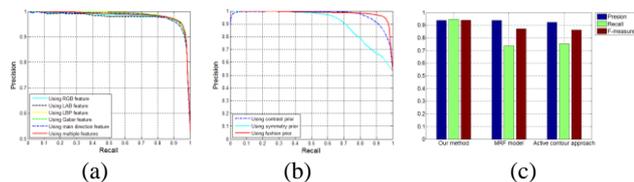


Figure 7. Precision–recall curves: (a) with different features and (b) with different priors, (c) overall precision, recall and F-measure values of the compared methods on 20 test images.

Finally, the formulas proposed by Lin and Nevatia (1998) for branch factor (BF) and detection percentage (DP) are used to compute the detection performance quantitatively. BF describes the ratio of incorrectly detected objects, and DP reflects the number of man-made objects in the image detected by the extraction algorithm. The objective is to maximize the DP and minimize the BF simultaneously.

$$BF = \frac{FP}{TP} \quad (9)$$

$$DP = \frac{100TP}{TP+FN} \quad (10)$$

where TP denotes the true positive or the number of man-made objects detected manually and automatically, FP represents the false positive or the number of false alarms and FN denotes the false negative or the number of undetected man-made objects.

The proposed method is evaluated on 20 test images in terms of BF and DP. The overall BF of the proposed method is 0.068, and its DP is 94.8%. These results are better than those of the methods by Zong et al. (2015) (i.e. BF, 0.095; DP, 73.7%) and Wang et al. (2009) (i.e. BF, 0.133; DP, 75.4%).

5. CONCLUSION

We have proposed a top-down approach to extract man-made objects from remote sensing images automatically by manifold ranking using a graph, which integrates colour, texture and main direction cues. As a kind of optimization algorithm, symmetry and contrast features are considered to obtain a priori value. We have evaluated the proposed algorithm on 20 remote sensing images. The proposed approach has presented decent overall quality. In our future work, we intend to use the more characteristics of man-made objects and develop an improved method for fusion of multiple features.

ACKNOWLEDGEMENTS

This work was supported by the High Resolution Images Surveying and Mapping Application System of China (Project No. AH1601-11)

REFERENCES

Achanta, R., Hemami, S., Estrada, F., and Susstrunk, S., 2009. Frequency-tuned salient region detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 22(9), pp. 1597-1604.

Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Susstrunk, S., 2010. SLIC superpixels. *EPFL*.

Blum, H., 1967. A transformation for extracting new descriptors of shape. *Models for the Perception of Speech & Visual Form*, 19(1), pp. 362-380.

Cai, F., Chen, H., and Ma, J., 2011. Man-made Object Detection Based on Texture Clustering and Geometric Structure Feature Extracting. *International Journal of Information Technology and Computer Science*, 3(2), pp. 9.

Felzenszwalb, P. F., and Huttenlocher, D. P., 2004. Efficient Graph-Based Image Segmentation. *International Journal of Computer Vision*, 59(2), pp. 167-181.

Florczyk, A. J., Ferri, S., Syrris, V., and Kemper, T., 2016. A New European Settlement Map From Optical Remotely Sensed Data. *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*, 9(5), pp. 1978-1992.

Gou, S. P., 2008. Image Segmentation of Quantum Evolutionary Programming-based Kernel Clustering Algorithm. *Computer Science*, 35(7), pp. 213-215.

Hinz, S., and Baumgartner, A., 2003. Automatic extraction of urban road networks from multi-view aerial imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 58(1), pp. 83-98.

Ji, S. P., and Yuan, X. X., 2007. A Method for Shadow Detection and Change Detection of Man-made Objects. *Journal of Remote Sensing*, 11(3), pp. 323-329.

Kapur, J. N., Sahoo, P. K., and Wong, A. K. C., 1980. A new method for gray-level picture thresholding using the entropy of the histogram. *Computer Vision Graphics & Image Processing*, 29(3), pp. 273-285.

Levinshstein, A., Dickinson, S., and Sminchisescu, C., 2013. Multiscale symmetric part detection and grouping. *International Journal of Computer Vision*, 104(2), pp. 117-134.

Levitt, S., and Aghdasi, F., 1997. Texture measures for building recognition in aerial photographs. *Communications and Signal Processing, 1997 South African Symposium on IEEE*, pp. 75-80.

Li, J., Najmi, A., and Gray, R. M., 2000. Image classification by a two-dimensional hidden Markov model. *Signal Processing IEEE Transactions*, 48(2), pp. 517-533.

Lin, C., and Nevatia, R., 1998. Building Detection and Description from a Single Intensity Image. *Computer Vision and Image Understanding*, 72(2), pp. 101-121.

Makantasis, K., Karantzas, K., Doulamis, A., and Loupos, K., 2015. Deep Learning-Based Man-Made Object Detection from Hyperspectral Data. *International Symposium on Visual Computing. Springer International Publishing*, 9474(1), pp. 717-727.

Reno, A. L., and Booth, D. M., 1999. Using models to recognise man-made objects. *Visual Surveillance, 1999. Second IEEE Workshop*, pp. 33-40.

Shi, J. and Malik, J., 2000. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8), pp. 888-905.

Turker, M., and Sumer, E., 2007. Building-based damage detection due to earthquake using the watershed segmentation of the post-event aerial images. *International Journal of Remote Sensing*, 29(11), pp. 3073-3089.

Wang, L., He, L., Mishra, A., and Li, C., 2009. Active contours driven by local Gaussian distribution fitting energy. *Signal Processing*, 89(12), pp. 2435-2447.

Wang, W., and Yang, X., 2010. Rapid, man-made object morphological segmentation for aerial images using a multi-scaled, geometric image analysis. *Image & Vision Computing*, 28(4), pp. 626-633.

Yang, C., Zhang, L. H., Lu, H. C., Ruan, X., and Yang, M. H., 2013. Saliency detection via graph-based manifold ranking. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 9(4), pp.3166-3173.

Zhou, D. Y., Weston, J., Gretton, A., Bousquet, O., and Scholkopf, B., 2003. Ranking on Data Manifolds. *Advances in Neural Information Processing Systems*, pp. 169-176.

Zong, K., Sowmya, A., and Trinder, J., 2015. Building Change Detection Based on Markov Random Field: Exploiting Both Pixel and Corner Features. *International Conference on Digital Image Computing: Techniques and Applications IEEE*, pp. 1-7.