2014 Third International Workshop on Earth Observation and Remote Sensing Applications

Sparse coding based airport detection from medium resolution Landsat-7 satellite remote sensing images

Gong Cheng, Junwei Han, Peicheng Zhou, Xiwen Yao, Dingwen Zhang, Lei Guo
School of Automation, Northwestern Polytechnical University, Xi'an, 710072, China
E-mail: gcheng@nwpu.edu.cn, junweihan2010@gmail.com

***Abstract*—A simple but effective method for airport detection from medium resolution Landsat-7 satellite remote sensing images based on sparse coding is presented. It consists of three phases: dictionary construction, sparse coding, and airport detection. Firstly, an over-complete dictionary is constructed using a set of airport training samples. Secondly, test images are scanned using multi-scale windows and each scanned window is sparsely coded in terms of atoms of the dictionary. Finally, sparsity concentration index of each scanned window is calculated based on the coding coefficients, which is used to decide the airport detection. Evaluations on publically available satellite images and comparisons with state-of-the-art approaches have demonstrated the superiority of the presented work.**

Keywords-airport detection; Landsat-7 satellite images; sparse coding; multi-scale

I. INTRODUCTION

The development of robust and reliable methods for automatic detection of meaningful targets from remote sensing images has been a profound challenge for computer vision community. As airport is an important civil and military target, the topic of automatic detection of airport has attracted increasing attention [1-8]. At the same time, it is still a difficult task to obtain better detection performance due to the difficulties such as airport appearance variation, cluttered background, and relatively low resolution of satellite images.

Most previous works mainly focused on recognizing runways for airport detection based on the premise that runways are the most prominent characteristics [1-3]. These methods adopted a heuristic or ad hoc manner and the success of them largely relies on the capability of extracting runway cues. In recent years, many airport detection approaches utilized supervised learning techniques, which treat target detection as a classification issue [4-9]. These classification-based methods basically consist of training stage and test stage. In training stage, a variety of airport samples are selected to train a classifier. In the test stage, each location in test images is compared against the classifier to decide if this location contains the target or not. With the help of prior information obtained from training samples, most of the classification-based methods have achieved good detection performance for high spatial resolution images. However, although existing approaches have developed some powerful features such as line-derived features [2, 3, 8], shape-derived features [5], and texture-derived features [4, 7], these features are generally appropriate for high resolution remote sensing images. A practical issue we have to consider is to detect airport from medium resolution remote sensing images, such as Landsat-7 satellite images. In this situation, the runways and airports may be incomplete and even blurred, which largely increases the difficulty of extracting discriminative and reliable features to characterize airports.

Recently, the sparse coding theory has been extensively studied and successfully applied to image restoration [10], face recognition [11], image classification [12, 13], and hyperspectral imagery analysis [14]. Its fundamental idea is to represent a high-dimensional original signal on a low-dimensional manifold in terms of a sparse linear combination of an over-complete bases or dictionary [15]. Inspired by the success of sparse coding in visual recognition tasks, we present a simple but effective method for airport detection from medium resolution Landsat-7 satellite remote sensing images using sparse coding theory. In contrast with line-based, shape-based or texture-based features that are more appropriate for high spatial resolution remote sensing images, our sparse coding-based approach directly uses the image patches as the features, which are more reliable for remote sensing images with medium or relatively low spatial resolution. Moreover, in sparse coding framework, the remote sensing images are sparsely coded by a few representative atoms in a low-dimensional manifold, which achieves an efficient computation.

II. SPARSE CODING THEORY

Sparse coding theory is originally proposed by [16] and [15], and then fully developed by [17] and [18]. Specifically, the sparse coding can be formulated as [11, 17, 18]:

$$\boldsymbol{y} = \mathbf{D}\boldsymbol{\alpha} = \alpha_1 \boldsymbol{d}_1 + \alpha_2 \boldsymbol{d}_2 + \cdots + \alpha_k \boldsymbol{d}_k , \tag{1}$$

where $\boldsymbol{y} = [y_1, y_2, \ldots, y_m]^T$ is an input signal, $\mathbf{D} = [\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_k]$ is an over-complete dictionary with each column $\boldsymbol{d}_i$ representing an atom, and $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \ldots, \alpha_k]^T$ is a coefficient vector. In general, the number of atoms k is large, whereas the signal dimension m is relatively small. To prevent $\mathbf{D}$ from being arbitrarily large, a common constrain that ℓ_2 norm of $(\boldsymbol{d}_i)_{i=1}^k$ is less than or equals to 1 is put, i.e.,

$$\sqrt{\boldsymbol{d}_i^T \boldsymbol{d}_i} \le 1, \quad i = 1, \ldots, k . \tag{2}$$

978-1-4799-4184-1/14/$31.00 ©2014 IEEE

Since real remote sensing images are usually noisy owing to atmospheric radiation and sensor noise, it is usually difficult to represent the test image exactly as a sparse superposition of the training samples. Similar to the work of [11], (1) can be modified to explicitly account for small dense noise by

$$\boldsymbol{y} = \mathbf{D}\boldsymbol{\alpha} + \boldsymbol{z}\,, \tag{3}$$

where $\boldsymbol{z}$ denotes the noise. Thus, the sparse solution can be approximately solved by using the following ℓ_1 norm minimization:

$$\hat{\boldsymbol{\alpha}} = \arg\min \|\boldsymbol{\alpha}\|_1 \text{ subject to } \|\boldsymbol{y} - \mathbf{D}\boldsymbol{\alpha}\|_2 \le \varepsilon\,, \tag{4}$$

where $\|\cdot\|_1$ denotes the ℓ_1 norm and $\|\cdot\|_2$ denotes the ℓ_2 norm. $\varepsilon \ge 0$ is an error tolerance, which determines the degree how well the representation approximates a test sample. In this paper, we utilize LARS–Lasso method [19] to solve (4).

III. AIRPORT DETECTION USING SPARSE CODING

As shown in Fig. 1, the presented airport detection method consists of three stages. The first stage is to construct an over-complete airport dictionary, which is done by using the pixel intensities of the down-sampled training samples containing airport targets. In the second stage, all test images are scanned using multi-scale windows. Each scanned window is smoothed, down-sampled and sparsely coded in terms of atoms from the dictionary. Finally, sparsity concentration index (SCI) for each scanned window is calculated based on the coding coefficients, which is used to decide if its corresponding window contains the airport or not. A non-maximum suppression scheme [20] is applied to eliminate redundant detections.

A. Dictionary construction

When constructing a dictionary for sparse coding, there are two basic requirements: (1) it should contain a wide range of types of atoms that can cover different structures of signals, and (2) it is over-complete with sufficient number of atoms. To meet the requirements and ensure the dictionary is robust to airport appearance and orientation variations, we manually cropped n sub-images from a remote sensing image dataset and each sub-image contains a representative type of airports with different appearance. We then rotate each sub-image in the step of 10° from 0° to 360°. Finally, we use the bounding boxes to label all airports in those sub-images, which construct the training samples. In this way, we can totally obtain $36n$ training samples.

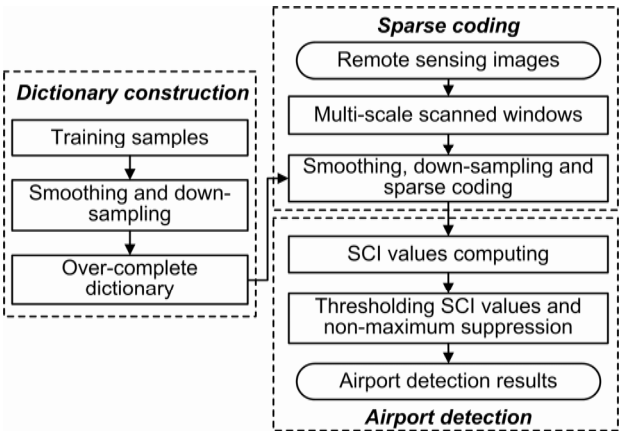

Figure 1. Block diagram of the presented approach. It consists of three stages: dictionary construction, sparse coding, and airport detection.

We next smooth all training samples using a Gaussian low-pass filter and down-sample them to the size of $a \times a$ where a is empirically set in the scope of [15, 25]. In our implementation, we set $a = 19$. Then, the columns of each down-sampled training sample are concatenated on top of each other to construct a a^2-dimensional feature vector, which is an atom of the dictionary. In this way, we can construct a dictionary $\mathbf{D} = [\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_n]$ with n sub-dictionary, where $\mathbf{D}_j = [d_{j,1}, d_{j,2}, \ldots, d_{j,36}]$, which is a sub-dictionary corresponding to the jth airport type and $j = 1, \ldots, n$. Hence, the size of dictionary $\mathbf{D}$ is $a^2 \times 36n$. To guarantee the dictionary is over-complete, a^2 is less than $36n$.

B. Sparse coding

To deal with the problem of target scale variation, we adopt multi-scale windows to scan each test image, and then represent each scanned window through sparse coding based on the constructed dictionary. Specifically, given a test remote sensing image, it is firstly scanned using multi-scale windows in the step of $\lceil s_{\mathrm{w}}/10 \rceil$ along the horizontal and vertical directions, where s_{w} is the size of each scanned window. The determination of the number of scanned windows depends on the requirement of detection accuracy and computational cost. If more scanned windows are used, we can achieve the better detection accuracy whereas the computational cost is expensive and vice versa. Our current implementation adopted 6 scanned windows empirically, which are in the sizes of $\lceil 5000/s_{\mathrm{r}} \rceil$, $\lceil 4500/s_{\mathrm{r}} \rceil$, $\lceil 4000/s_{\mathrm{r}} \rceil$, $\lceil 3500/s_{\mathrm{r}} \rceil$, $\lceil 3000/s_{\mathrm{r}} \rceil$, and $\lceil 2500/s_{\mathrm{r}} \rceil$ respectively. Here s_{r} is the spatial resolution of test image. Then, for each scanned window, we smooth it using a Gaussian low-pass filter, down-sample it to the size of $a \times a$, and then convert it to a a^2-dimensional feature vector. Finally, each scanned window is linearly coded over the dictionary of atoms by using (4) to obtain the coding coefficient vector $\hat{\boldsymbol{\alpha}}$.

C. Airport detection

Once the coefficient vector $\hat{\boldsymbol{\alpha}}$ for each scanned window is obtained, its SCI value can be defined as [11]:

$$\mathrm{SCI}(\hat{\boldsymbol{\alpha}}) = \frac{n \cdot \max_j \left\| \delta_j(\hat{\boldsymbol{\alpha}}) \right\|_1 / \left\| \hat{\boldsymbol{\alpha}} \right\|_1 - 1}{n - 1} \in [0,1]\,, \tag{5}$$

where $\delta_j(\hat{\boldsymbol{\alpha}}) = [0, \ldots, 0, \alpha_{j,1}, \alpha_{j,2}, \ldots, \alpha_{j,36}, 0, \ldots, 0]^T \, (j = 1, \ldots, n)$ is a new vector whose entries are zero except those associated with the jth airport type. For a solution $\hat{\boldsymbol{\alpha}}$, if $\mathrm{SCI}(\hat{\boldsymbol{\alpha}}) = 1$, the

978-1-4799-4184-1/14/$31.00 ©2014 IEEE

test image patch is represented using only samples from a single airport type, and if $\mathrm{SCI}(\hat{\boldsymbol{\alpha}}) = 0$, it means the sparse coefficients are spread evenly over all airport types.

We decide a test image patch labeled by a scanned window to contain an airport target if $\mathrm{SCI}(\hat{\boldsymbol{\alpha}}) \geq \tau$, where $\tau \in (0,1)$ is a threshold measuring how confident that a test image patch contains an airport. Fig. 2 illustrates the sparse coefficients of an airport and a non-airport test image patch. Their SCI values are 0.90 and 0.27. In Fig. 2(a)–(b) and (d)–(e), the original image patches and the down-sampled results are shown. The sparse coefficients of Fig. 2(a) and (d) are shown in Fig. 2(c) and (f). It can be seen from the results that the coefficient vector of airport is sparse including a small number of nonzero entries whereas the coefficient vector of non-airport is very dense including a large number of nonzero entries.

A practical problem needs to be considered in airport detection. Normally, a couple of image patches, with same or different scanned window sizes, near an airport are all likely to be detected as the airport targets by using the above described algorithm, which results in a few false positives. We adopt a simple and effective post-processing procedure, i.e. non-maximum suppression scheme [20] to address this problem. Specifically, we sort all scanned windows in descending order based on their corresponding SCI values and greedily select the windows with the highest SCI values. The windows which are at least 50% covered by any previously selected windows are eliminated.

IV. Experimental Results and Discussions

We comprehensively evaluate the presented approach using medium-spatial-resolution single-band remote sensing images of Landsat-7 Enhanced Thematic Mapper Plus (ETM+) that is publically available at http://datamirror.csdb.cn/index.jsp. In our evaluations, 55 30-m-spatial-resolution mid-infrared (MIR) images (band 7; wavelength: 2.08–2.35 micron) with the size of 8320×7400 and 26 15-m-spatial-resolution panchromatic images (band 8; wavelength: 0.52–0.90 micron) with the size of 16640×14800 were adopted. These 81 images were taken with a time spanning from December 9, 1999 to March 19, 2003 and have gone through radiometric and geometric calibration. Comparing with 1-m-spatial-resolution IKONOS images used by [6], our data can be considered as medium spatial resolution. In our experiment, 16 MIR images and 9 panchromatic images were selected as the training data and the rest 56 images were used for testing. We manually labeled 136 airports in test images used as ground truth and labeled 30 airports with different appearances from 25 training images to construct a 361×1080 dictionary offline.

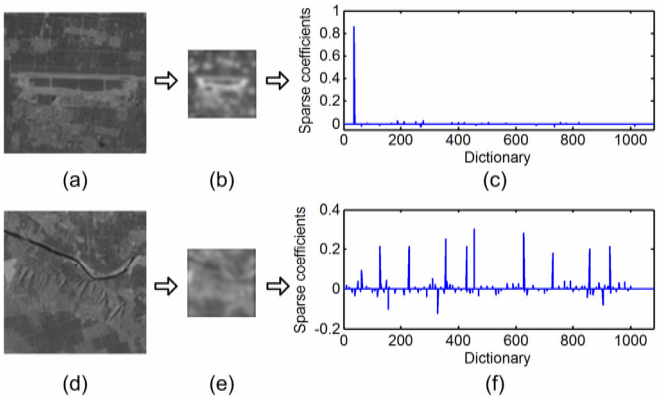

Figure 2. Sparse coefficients of an airport and a non-airport test image patch. (a) Original image patch of an airport. (b) Smoothed and down-sampled result of image patch (a). (c) Sparse coefficients of image patch (a). (d) Original image patch of a non-airport. (e) Smoothed and down-sampled result of image patch (d). (f) Sparse coefficients of image patch (d).

A. Evaluation criterion

Similar to [21] and [20], we adopted the standard Precision-Recall curve (PRC) to quantitatively evaluate the airport detection method, which plots *Recall* on the x-axis and *Precision* on the y-axis. *Recall* measures the fraction of positives that are correctly identified and *Precision* measures the fraction of detections that are true positives. Let N_{c}, N_{f}, and N_{t} denote the number of true positives, the number of false positives, and the number of total airport targets in test images. The *Recall* and *Precision* can be formulated as:

$$Recall = N_{\mathrm{c}} / N_{\mathrm{t}}\,, \tag{6}$$

$$Precision = N_{\mathrm{c}} / (N_{\mathrm{c}} + N_{\mathrm{f}})\,. \tag{7}$$

In addition, we also adopted the Average Precision (AP) [22] to quantitatively evaluate the performance of an object detection system, which is a standard metric used by PASCAL Visual Object Classes (VOC) challenge and can be obtained by computing the area under the PRC. The higher the AP value is, the better the performance and vice versa.

The correct detection is defined as follows. For each airport target, we manually label its ground truth bounding box and estimate its central location $(i_{\mathrm{true}}, j_{\mathrm{ture}})$ in the original test image. We consider a detection with central location $(i_{\mathrm{det}}, j_{\mathrm{det}})$ to be correct if the following two conditions are satisfied: (1) $|i_{\mathrm{true}} - i_{\mathrm{det}}| \leq \rho$, $|j_{\mathrm{true}} - j_{\mathrm{det}}| \leq \rho$, and (2) the bounding boxes at the detected and true locations have an overlap of at least θ_{area}. Here, the parameters ρ and θ_{area} are chosen depending on the size of ground truth bounding box. In our experiment, we set $\rho = \lceil 3b/10 \rceil$ and $\theta_{\mathrm{area}} = 70\%$, where b is the dimension of ground truth bounding box. In addition, if two or more scanned windows satisfy these conditions at the same location, only one is treated as the correct detection.

B. Experimental results

The presented approach was evaluated on 39 8320×7400 MIR images and 17 16640×14800 panchromatic images containing 136 airports. In the presented method, the error tolerance ε is a critical parameter. Our first experiment is to evaluate how the parameter affects the detection performance. Fig. 3 shows the PRCs and their AP values when varying ε from 0.30 to 0.50. According to our experiment, the parameter influences the detection results moderately and the best

978-1-4799-4184-1/14/$31.00 ©2014 IEEE

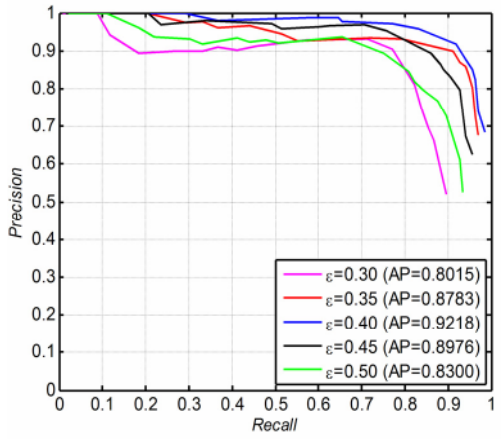

Figure 3. Detection performance of different error tolerances ε in terms of PRC and AP.

performance was achieved when $\varepsilon = 0.4$. Consequently, we empirically set $\varepsilon = 0.4$ in our all evaluations.

We implemented the presented algorithm using MATLAB R2010b. A fixed set of parameter values were utilized for all test images. On a 24-core Lenovo Server with Intel Xeon CPU of 2.8 GHz, for a remote sensing image with the resolution of 8320×7400 or 16640×14800, the sparse coding costs about 133 seconds and the airport detection can be accomplished within one second. The processing time is reasonable considering the large size of test images.

Fig. 4 shows some examples of the airport detection results using the presented method where the true positives, false negatives, and false positives are highlighted by blue, red and yellow rectangles, respectively. Their locations and spatial resolutions are denoted below each example. These examples were selected from our test dataset. As can be seen from Fig. 4, the presented approach has successfully detected and located most airports with different sizes in our test images. However, our approach failed to detect one airport in Fig. 4(d) because the contrast between the airport and the background is quite low. It is even difficult for human vision system to distinguish that airport from background. There is a false positive which is actually a building shown in Fig. 4(e) due to the fact that the building is quite similar to airports in structure and shape. Since our approach mainly depends on the patch-based spatial and structure information, the false detection results in Fig. 4(d)–(e) are reasonable.

The quantitative evaluation results are shown in Table 1. As can be seen from Table 1, the value of threshold τ considerably affects the detection performance. There is a trade-off between the number of correct detections and number of false detections. A low threshold can achieve a good *Recall* but a poor *Precision* and vice versa. The optimal threshold is considered to have the highest $F1\text{-}measure$:

$$F1\text{-}measure = 2 \cdot Recall \cdot Precision / (Recall + Precision). \quad (8)$$

In our experiment, we obtained the optimal trade-off when $\tau = 0.75$, which corresponds to 91.91% *Recall* and 91.91% *Precision*.

Table I. Detection results of our method by using different thresholds

Threshold τ	N_c	*Recall*	N_f	*Precision*
0.65	130	95.59%	23	84.97%
0.70	128	94.12%	18	87.67%
0.75	125	91.91%	11	91.91%
0.80	121	88.97%	9	93.08%
0.85	113	83.09%	5	95.76%
0.90	105	77.21%	3	97.22%

C. *Comparison with previous work*

We compared our approach with Tao's method [6] and Qu's method [4] on the same dataset. The parameters of Tao's method were set in terms of the instructions provided by [6]. In our experiments, we set $R = 140$ and $S = 4$. Here R indicates the maximum cluster diameter and S is the minimum number of Scale-invariant feature transform (SIFT) key points in one cluster. The comparison results are shown in Fig. 5. As can be seen from Fig.5, the presented approach works better than Tao's method [6] and Qu's method [4].

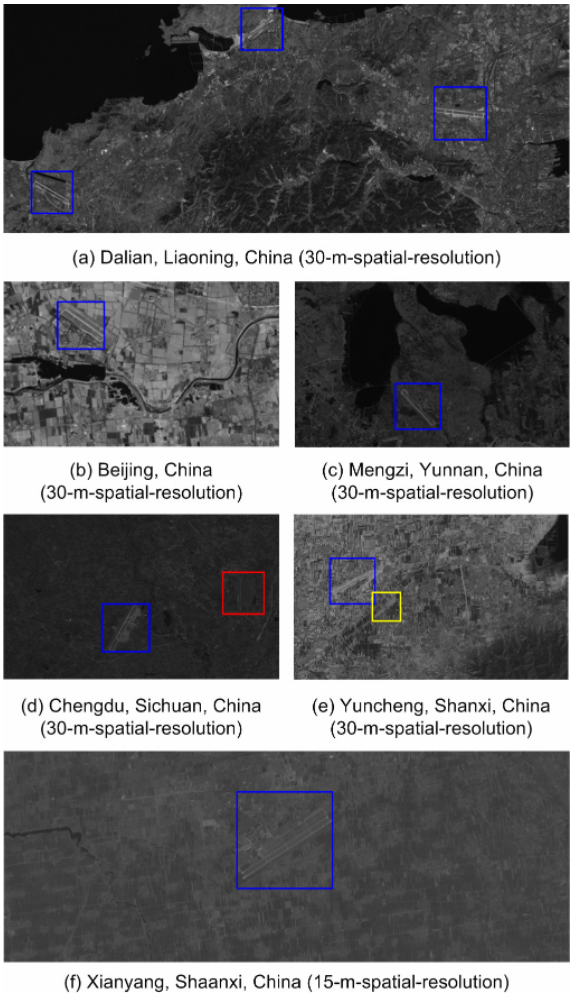

Figure 4. Examples of airport detection.

978-1-4799-4184-1/14/$31.00 ©2014 IEEE

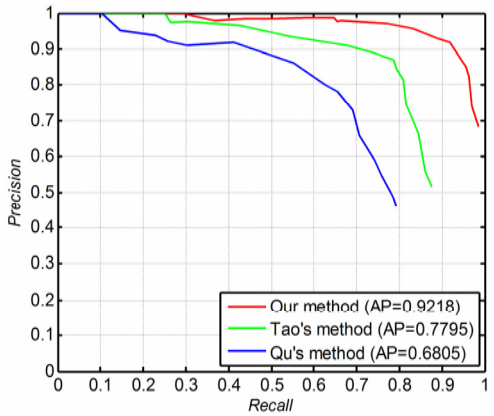

Figure 5. Performance comparisons of our airport detection method and the methods proposed by [6] and [4].

The superior comparison results of our sparse coding-based airport detection method can be easily explained. The detection performance of Tao's method [6] and Qu's method [4] significantly relies on their extracted line-based or key points-based features. These features are not robust and reliable to cope with the problems such as appearance variation, occlusion, and medium resolution of remote sensing images. In contrast, in the presented method, the choice of feature becomes less important. Image patches are directly utilized as the base for training and test, which are relatively less sensitive to medium resolution remote sensing images.

V. CONCLUSIONS

We presented a simple yet effective approach to detect airports from medium resolution single-band Landsat-7 satellite images based on sparse coding theory. In contrast with line-based, shape-based or texture-based features that are more appropriate for images with high spatial resolution, our sparse coding-based approach directly uses the image patches as the features, which are more reliable for medium-spatial-resolution remote sensing images. Evaluations on publically available Landsat-7 satellite remote sensing images dataset and comparisons against previous work have demonstrated the effectiveness of the presented work. Our future work mainly include: (1) integrating spectral information and contextual cues with the presented method to help reduce false positives and improve detection performance; (2) extending the presented framework to multi-class object detection in remote sensing images by constructing a class-specific dictionary.

ACKNOWLEDGMENT

This work was supported in part by NSFC under Grants 61005018 and 91120005, Program for New Century Excellent Talents in University under grant NCET-10-0079, and China Postdoctoral Science Foundation under Grant 2014M552491.

REFERENCES

[1] J. Han, L. Guo, and Y. Bao, "A method of automatic finding airport runways in aerial images," in *Proc. ICSP*, 2002, pp. 731-734.

[2] Y. Pi, L. Fan, and X. Yang, "Airport detection and runway recognition in SAR images," in *Proc. IGARSS*, 2003, pp. 4007-4009.

[3] D. Liu, L. He, and L. Carin, "Airport detection in large aerial optical imagery," in *Proc. ICASSP*, 2004, pp. 761-764.

[4] Y. Qu, C. Li, and N. Zheng, "Airport detection base on support vector machine from a single image," in *Proc. ICICS*, 2005, pp. 546-549.

[5] S. Zhang, Y. Lin, X. Zhang, and Y. Chen, "Airport automatic detection in large space-borne SAR imagery," *J. Syst. Eng. Electron.,* vol. 21, no. 3, pp. 390-396, 2010.

[6] C. Tao, Y. Tan, H. Cai, and J. Tian, "Airport detection from large IKONOS images using clustered SIFT keypoints and region information," *IEEE Geosci. Remote Sens. Lett.,* vol. 8, no. 1, pp. 128-132, 2011.

[7] Ö. Aytekin, U. Zöngür, and U. Halici, "Texture-based airport runway detection," *IEEE Geosci. Remote Sens. Lett.,* vol. 10, no. 3, pp. 471-475, 2013.

[8] X. Wang, Q. Lv, B. Wang, and L. Zhang, "Airport detection in remote sensing images: a method based on saliency map," *Cognitive Neurodynamics,* vol. 7, no. 2, pp. 143-154, 2013.

[9] G. Cheng, J. Han, L. Guo, X. Qian, P. Zhou, X. Yao, and X. Hu, "Object detection in remote sensing imagery using a discriminatively trained mixture model," *ISPRS J. Photogramm. Remote Sens.,* vol. 85, pp. 32-43, 2013.

[10] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Trans. Image Process.,* vol. 17, no. 1, pp. 53-69, 2008.

[11] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 31, no. 2, pp. 210-227, 2009.

[12] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *Proc. CVPR*, 2009, pp. 1794-1801.

[13] B. J. Culpepper, J. Sohl-Dickstein, and B. A. Olshausen, "Building a better probabilistic model of images by factorization," in *Proc. ICCV*, 2011, pp. 2011-2017.

[14] A. Castrodad, Z. Xing, J. B. Greer, E. Bosch, L. Carin, and G. Sapiro, "Learning discriminative sparse representations for modeling, source separation, and mapping of hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.,* vol. 49, no. 11, pp. 4263-4281, 2011.

[15] B. Olshausen, and D. Field, "Sparse coding with an overcomplete basis set: a strategy employed by V1?," *Vision Research,* vol. 37, no. 23, pp. 3311-3325, 1997.

[16] S. G. Mallat, and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Image Process.,* vol. 41, no. 12, pp. 3397-3415, 1993.

[17] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. computing,* vol. 20, no. 1, pp. 33-61, 1999.

[18] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution," *Comm. Pure and Applied Math.,* vol. 59, no. 6, pp. 797-829, 2006.

[19] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, "Least angle regression," *The Annals of statistics,* vol. 32, no. 2, pp. 407-499, 2004.

[20] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 32, no. 9, pp. 1627-1645, 2010.

[21] S. Agarwal, and D. Roth, "Learning a sparse representation for object detection," in *Proc. ECCV*, 2002, pp. 113-130.

[22] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *Int. J. Comput. Vis.,* vol. 88, no. 2, pp. 303-338, 2010.

978-1-4799-4184-1/14/$31.00 ©2014 IEEE