Feature Extraction by Rotation-Invariant Matrix Representation for Object Detection in Aerial Image

Guoli Wang, Xinchao Wang, Bin Fan, and Chunhong Pan

***Abstract*— This letter proposes a novel rotation-invariant feature for object detection in optical remote sensing images. Different from previous rotation-invariant features, the proposed rotation-invariant matrix (RIM) can incorporate partial angular spatial information in addition to radial spatial information. Moreover, it can be further calculated between different rings for a redundant representation of the spatial layout. Based on the RIM, we further propose an RIM_FV_RPP feature for object detection. For an image region, we first densely extract RIM features from overlapping blocks; then, these RIM features are encoded into Fisher vectors; finally, a pyramid pooling strategy that hierarchically accumulates Fisher vectors in ring subregions is used to encode richer spatial information while maintaining rotation invariance. Both of the RIM and RIM_FV_RPP are rotation invariant. Experiments on airplane and car detection in optical remote sensing images demonstrate the superiority of our feature to the state of the art.**

***Index Terms*— Feature extraction, Fisher vector, object detection, ring pyramid pooling (RPP), rotation-invariant matrix (RIM).**

I. Introduction

FEATURE extraction is an important procedure in many visual tasks [1]–[4]. In optical remote sensing images, detecting objects with arbitrary orientations is a major challenge. To handle this challenge, the most direct way is to augment training set by rotating objects with multiple orientations so as to train a rotation-invariant classifier [5]. However, due to the loss of spatial information, this method has a complicated feature space and thus increases the difficulty in learning a discriminative classifier.

To alleviate this problem, Zhang *et al.* [1] extended the histograms of oriented gradients (HOGs) [6] by aligning an image region with an estimated dominant orientation. Similarly, Chen *et al.* [7] proposed to input an rotated image into a deep belief net by computing the main axis of an aircraft. However, the dominant orientation estimation is not always accurate enough, so the stability of the extracted features is not guaranteed.

Lazebnik *et al.* [8] proposed to construct rotation-invariant features by accumulating histograms of intensities or radial gradient orientations (RGOs) in several concentric rings. However, their ring-shaped pooling strategy only incorporates radial spatial information while discarding angular spatial information. As a result, the obtained features are not discriminative enough. Lin *et al.* [9] proposed radial gradient angle to describe the rotation-invariant attribute of each pixel by computing the difference between the radial angle and the dominant gradient orientation around each pixel. However, it lacks structural information and thus has poor discrimination.

To encode richer spatial information while holding the rotation invariance, we propose a novel feature extraction method. The contribution lies in three aspects: 1) we propose the rotation-invariant matrix (RIM) to incorporate partial angular spatial information while being rotation invariant; 2) Fisher vector [10] is used to encode the RIM feature for a stronger discriminative ability; and 3) finally, we hierarchically accumulate Fisher vectors in a ring pyramid schema to incorporate more spatial information while maintaining the rotation invariance. These aborative designs help to improve the discriminative ability of the proposed feature. The experimental evaluations on airplane and car detection in optical remote sensing images demonstrate its effectiveness.

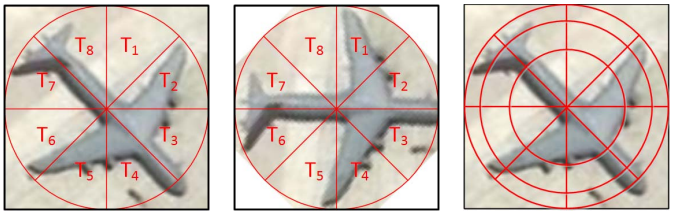


Fig. 1. Spatial division used for calculating the proposed RIM, (a) an image is divided into 8 sectors, (b) is a rotated image version of (a), (c) is an illustration of dividing the image region into several concentric rings and each ring is further divided into several sectors.

II. Rotation-Invariant Feature Construction

A. Rotation-Invariant Matrix

For an image region R, we divide it into s sectors $\{T_1, T_2, \ldots, T_s\}$, as shown in Fig. 1(a). Denoting

Manuscript received June 13, 2016; revised January 20, 2017; accepted February 21, 2017. Date of publication April 12, 2017; date of current version May 19, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 61403375, Grant 61472119, and Grant 91338202, and in part by the Priority Academic Program Development of Jiangsu Higher Education Institutions and the Jiangsu Collaborative Innovation Center on Atmospheric Environment and Equipment Technology. *(Corresponding author: Bin Fan.)*

G. Wang is with the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China, and also with the School of Computer and Control Engineering, University of Chinese Academy of Sciences, Beijing 101408, China (e-mail: glwang@nlpr.ia.ac.cn).

X. Wang is with the Image Formation and Processing Group, Beckman Institute, University of Illinois Urbana–Champaign, Champaign, IL 61801 USA (e-mail: xinchao@illinois.edu).

B. Fan and C. Pan are with the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: bfan@nlpr.ia.ac.cn; chpang@nlpr.ia.ac.cn).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LGRS.2017.2683495

$\{f_1, f_2, \ldots, f_s\}$ as the feature vectors computed from these sectors, then the image region R can be described by a feature matrix $M_R = [f_1^T; f_2^T; \ldots; f_s^T]$. In this way, when the image is rotated by $(2\pi/s)$ as shown in Fig. 1(b), the corresponding rotated region R' can be described by the feature matrix $M_{R'} = [f_1'^T; f_2'^T; \ldots; f_s'^T]$. Suppose the feature vector f_i computed from each sector is rotation invariant, then we have $f_1'^T = f_2^T$, $f_2'^T = f_3^T, \ldots, f_s'^T = f_1^T$. Apparently, $M_{R'}$ is the circular row permutation of M_R. Denoting P_{ij} as the permutation matrix which exchanges the ith and jth rows of a matrix, then $M_{R'} = P_{(s-1)s} \ldots P_{23} P_{12} M_R$. Since $P_{ij}^T P_{ij} = I$, where I denotes the identity matrix, the following equation holds:

$$\begin{aligned} M_{R'}^T M_{R'} &= (P_{(s-1)s} \ldots P_{12} M_R)^T (P_{(s-1)s} \ldots P_{12} M_R) \\ &= \left(M_R^T P_{12}^T \ldots P_{(s-1)s}^T\right)(P_{(s-1)s} \ldots P_{12} M_R) \\ &= M_R^T P_{12}^T \ldots \left(P_{(s-1)s}^T P_{(s-1)s}\right) \ldots P_{12} M_R \\ &= M_R^T M_R. \end{aligned} \tag{1}$$

It is clear that the rotation variance is eliminated by the matrix multiplication. Therefore, $M_R^T M_R$ is rotation invariant as long as the feature vector f_i of each sector is rotation invariant. We call $M_R^T M_R$ as RIM. As it is calculated on features from different sectors, it can incorporate angular spatial information to some extent while maintaining the rotation invariance, resulting in a high discrimination.

In this letter, two kinds of rotation-invariant features are used for each sector, although RIM is independent of the specific choice of rotation-invariant features. The first one accumulates a histogram of RGOs, which are calculated on the radial local coordinate systems [8]. As each local coordinate system is constructed independent of image rotation, the RGO is rotation invariant theoretically. The other one aggregates a histogram of radial gradient angles (RGAs) [9]. Along with image rotation, the gradient orientation and the radial angle will simultaneously change with the same angle, and hence, their difference, RGA, is invariant to in-plane rotation. Due to the space limit, please refer to [8] and [9] for more details about these two rotation invariants. In the following, we show how to effectively combine these two rotation-invariant histograms to acquire more mutual information.

To incorporate radial spatial information into the RIM, we consider to further divide the image region into several concentric rings [Fig. 1(c)] and extract multiple RIMs from these rings. Suppose the feature matrices based on RGA and RGO in the ith ring are MA_i and MO_i respectively, we can obtain various RIMs in the following ways:

$$\begin{aligned} &(a)\ \mathrm{MA}_i^T \mathrm{MA}_i \text{ and } \mathrm{MO}_i^T \mathrm{MO}_i \quad \forall i \\ &(b)\ \mathrm{MA}_i^T \mathrm{MA}_j \text{ and } \mathrm{MO}_i^T \mathrm{MO}_j \quad i \neq j \\ &(c)\ \mathrm{MA}_i^T \mathrm{MO}_j \quad \forall i, j. \end{aligned} \tag{2}$$

The final RIM feature collects all the elements in (2) with different rings. Note that the RIM in the first row of (2) is a semipositive definite symmetric matrix, and hence, we just use its upper triangular elements. The matrices in the second and the third rows provide redundant information of the spatial layout, which is useful in improving the descriptive ability of the RIM feature.

Algorithm 1 RIM Feature Construction

Input: Image Region R
Output: RIM Feature X
1: Divide R into several rings, which will be partitioned into several sectors again;
2: Calculate the histograms of RGA and RGO for each sector, then $L1$ normalize these histograms;
3: Construct the feature matrices MA_i *and* MO_i for each ring and calculate multiple RIMs according to Eq. (2);
4: $L2$ normalize the elements of each RIM and collect all of them as the final feature X.

It is worth to point out that there are some tricks in constructing the RIM feature: 1) when dividing the image region into rings, we try to keep the number of image pixels in each ring approximately identical; 2) similar to scale-invariant feature transform (SIFT), each gradient orientation contributes to its four neighboring sectors for histogram accumulation; 3) before constructing the feature matrix, $L1$ normalization is used for the histogram of each sector; and 4) $L2$ normalization is used for each RIM in (2). According to our experiments, all these four tricks help to improve its discriminative ability and robustness. Algorithm 1 gives the pseudocode of the procedure to extract the RIM feature.

RIM can be directly used to describe the whole object region; however, this low-level representation is not discriminative sufficiently. Therefore, we consider to use RIM features to densely describe overlapping blocks, which are further encoded into Fisher vectors [10] for stronger discrimination.

B. Rotation-Invariant Matrix With Fisher Vector

Suppose K Gaussian models are generated by the Gaussian mixture model from a training set of the principal component analysis (PCA) reduced RIM features, ω_k, μ_k, and σ_k^2, respectively, denote the mixture weight, mean vector, and variance vector of the kth Gaussian distribution (assuming that the covariance matrix is diagonal). For a reduced RIM feature x_i, its Fisher vector consists of the gradients of each Gaussian $\mathcal{G}_{\mu_k}^{x_i}$ and $\mathcal{G}_{\sigma_k}^{x_i}$ as follows:

$$\begin{aligned} \mathcal{G}_{\mu_k}^{x_i} &= \frac{\gamma_i(k)}{\sqrt{\omega_k}} \left(\frac{x_i - \mu_k}{\sigma_k}\right) \\ \mathcal{G}_{\sigma_k}^{x_i} &= \frac{\gamma_i(k)}{\sqrt{2\omega_k}} \left[\frac{(x_i - \mu_k)^2}{\sigma_k^2} - 1\right] \end{aligned} \tag{3}$$

where $\gamma_i(k) = ((\omega_k f_k(x_i; \mu_k, \sigma_k))/(\sum_{j=1}^K \omega_j f_j(x_i; \mu_j, \sigma_j)))$ is the probability of x_i belonging to the kth Gaussian. Suppose the dimension of x_i is D, then its Fisher vector will be a $2DK$-dimensional feature vector.

Fig. 2 shows the whole pipeline of our feature construction. First, the input image is densely partitioned into overlapping blocks and RIM features are extracted from these blocks; then applying PCA to these RIM features and the reduced features are subsequently encoded into Fisher vectors; finally, a ring

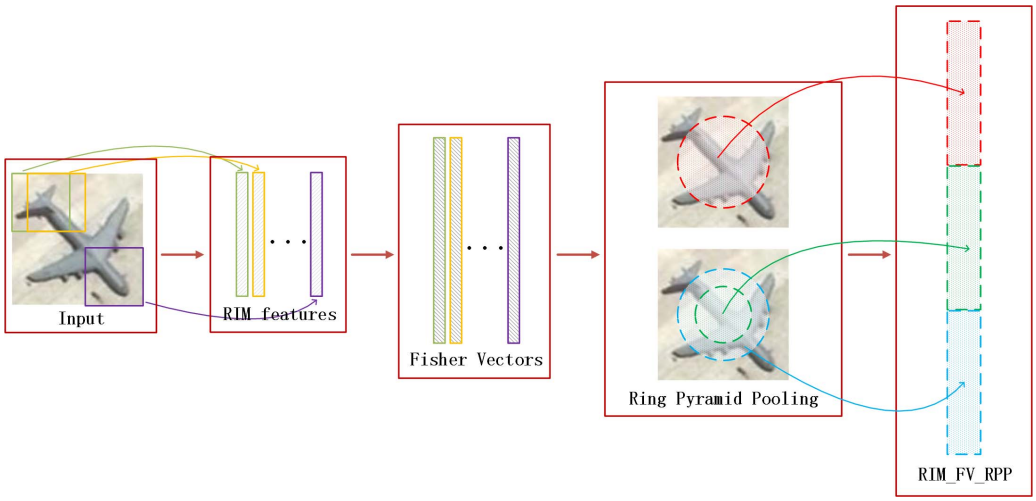


Fig. 2. Pipeline of constructing the proposed RIM_FV_RPP feature.

pyramid pooling (RPP) strategy dividing the image into ring subregions is used to accumulate Fisher vectors, respectively. Note that the final feature, which we call RIM_FV_RPP, is rotation invariant naturally due to the rotation invariance of RIM, Fisher Vector and RPP. Densely extracted RIM features could provide redundant structure information of the image region, and the incorporated Fisher vector leads to a higher discriminative ability; moreover, RPP encodes richer spatial information. Due to these facts, our feature has a very strong discriminative ability for rotated objects, which we will show in the following experiments.

III. Experiments

A. Experimental Setup

Two data sets are used to evaluate our method.

1) The airplane data set [11] consists of 50 training and 50 testing images (about 1000×800 pixels) downloaded from the Google Earth. To accelerate the detection speed, binarized normed gradients [12] are first used to generate airplane proposals. As the sizes of airplanes range from 700 to 25 488 pixels, three square windows in the sizes of $\{39 \times 39, 75 \times 75, 135 \times 135\}$ are used to generate proposals, which are subsequently resized to 32×32 for feature extraction.
2) The TAS data set [13], [14] includes 30 aerial images (about 792×636 pixels) with 1319 labeled cars. Since this data set provides proposals with a fixed size of 45×45, we directly use them for detection. As the experimental settings in [13], fivefold cross validation is performed. We report the average performance across all folds.

We extract a feature vector from each proposal and fit it into a classifier (linear support vector machine (SVM) is used in this letter) for object detection. When the intersection over union [15] between a proposal and an arbitrary groundtruth bounding box exceeds 0.5, the proposal is regraded as a positive; otherwise, it is treated as a negative. We consider to select hard negatives to train a discriminative classifier.

TABLE I
Parameter Settings of the Proposed Feature

r	s	b	bs	st	d	g	l
6	8	6	20	1	160	60	2

It first randomly selects partial training negatives for learning an original classifier; then, the classifier is applied to all training negatives and outputs their scores; subsequently, these negatives are sorted in a descending order by their scores and the top n negatives (i.e., hard negatives) are chosen to train the final classifier. n is set according to the original ratio between positives and negatives. For the airplane data set, n is $6N$, while for the car data set, n is $2N$, where N denotes the number of training positives. Finally, a nonmaximum suppression procedure [16] is used to remove overlapping proposals. We evaluate the detection performance by average precision (AP), which is widely used in object detection.

B. Parameter Evaluation

RIM_FV_RPP totally has eight parameters needed to be evaluated, including the number of rings r, sectors s and histogram bins b for RIM, the size bs of each block, the stride st of overlapping blocks, the reduced dimension d by PCA, the number of mixture Gaussians g, and the pyramid levels l. When evaluating these parameters, we only adjust one and keep the others at the best settings. The parameter evaluation is conducted on the mnist-rot data set [17], which is very different from the two data sets we used for performance evaluation. It contains 12 000 and 50 000 handwritten digit images for training and testing, respectively. Each image with a fixed size of 28×28 is sampled from MNIST and then randomly rotated.

The error rates with different parameter settings are shown in Fig. 3. As shown in Fig. 3(a)–(c), dividing the block into six rings with eight sectors and aggregating a histogram with six bins in each sector achieves the lowest error rate. Fig. 3(d) and (e) shows that extracting RIM features in large overlapping blocks efficiently decreases the error rate, and the

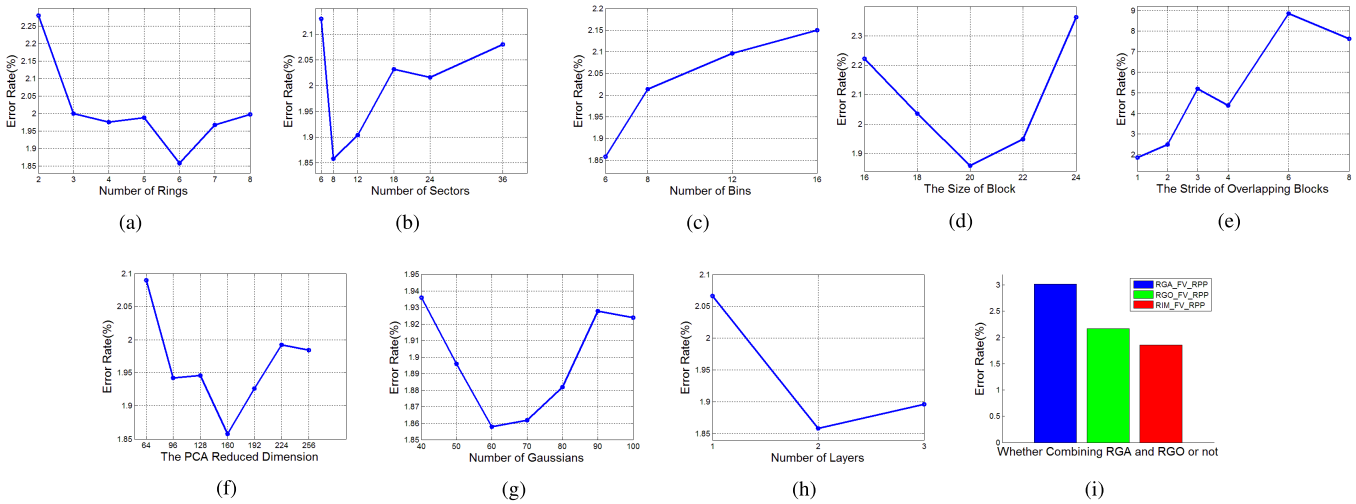


Fig. 3. Parameters evaluation on the mnist-rot data set. (a) Parameter r. (b) Parameter s. (c) Parameter b. (d) Parameter bs. (e) Parameter st. (f) Parameter d. (g) Parameter g. (h) Parameter l. (i) The influence of redundant information representation.

TABLE II
APs OF DIFFERENT METHODS ON AIRPLANE DETECTION

Method/Feature	AP(%)
HOG + linear SVM	34.16
Weakly supervised with LLC [11]	66.42
Transferred Deep Features + Negative Bootstrapping [4]	76.26
RIM + linear SVM	79.59
SIFT_FV_RPP + linear SVM	83.35
RIM_VLAD_RPP + linear SVM	85.36
RIM_FV_RPP + linear SVM	92.48

TABLE III
APs OF DIFFERENT METHODS ON THE TAS DATA SET

Method/Feature	AP(%)
HOG + linear SVM [22]	40.5
HOG + slot kernel structured SVM [22]	75.7
RC-RBM + Gradients IHOF + linear SVM [23]	77.6
Rotation-invariant HOG + Random Forest [24]	84.2
RIM + linear SVM	80.11
SIFT_FV_RPP + linear SVM	88.81
RIM_VLAD_RPP + linear SVM	89.40
RIM_FV_RPP + linear SVM	90.94

lowest error rate can be obtained at $bs = 20$ and $st = 1$. As shown in Fig. 3(f), the best performance is achieved at a dimension as low as $d = 160$. Fig. 3(g) shows that our method only needs a small number of Gaussians and setting $g = 60$ could get the lowest error rate. From Fig. 3(h), the RPP with $l = 2$ helps to further decrease the error rate. As a summary, Table I lists the final parameter settings, and we keep them consistent in the subsequent experiments. RGA_FV_RPP or RGO_FV_RPP shows the result of using RGA or RGO, respectively. The results in Fig. 3(i) demonstrate that combining these two rotation invariants could get better performance than using a single one, which further shows the redundant representation in (2) helps to improve the discriminative ability.

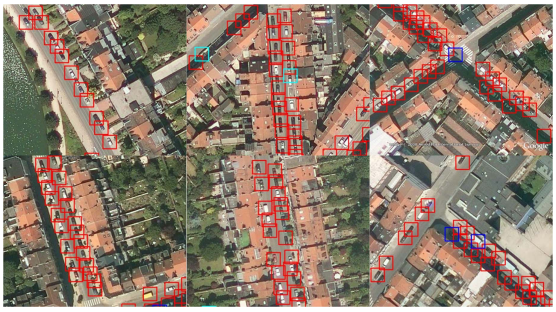
C. Performance Evaluation

Table II lists the results of different methods on airplane detection. HOG [6] performs the worst in all the methods. RIM performs better than HOG, increasing more than 45% AP. Once using RIM to form RIM_FV_RPP, a better performance can be achieved. By comparison, we also implement an SIFT_FV_RPP method by replacing RIM with SIFT [18]. As can be seen, RIM_FV_RPP achieves 9.13% higher AP than SIFT_FV_RPP. Besides Fisher vector, vectors of locally aggregated descriptors [19] are also a very popular method for incorporating low-level features. As a result, we implement a method called RIM_VLAD_RPP. It can be seen that RIM_FV_RPP performs better than RIM_VLAD_RPP. For all these evaluated features, they are implemented based on vlfeat [20]. Meanwhile, liblinear [21] is used as our SVM classifier. The best reported results of this data set are those from [4] and [11]; however, as the rotation invariance is ignored in these methods, their performance is even worse than using only RIM feature. The best result is obtained by RIM_FV_RPP, which obtains 92.48% AP.

Table III lists the results of different methods on car detection. HOG only achieves 40.5% AP with a linear SVM. Vedaldi *et al.* [22] learned the equivariant structured SVM to incorporate the rotation invariance, which improves AP to 75.7%. Schmidt and Roth [23] proposed a novel framework to learn transformation-invariant features. Liu *et al.* [24] builded rotation-invariant features in a Fourier space. Although all these methods [22]–[24] can deal with rotation problem, the higher performance on the TAS data set could be achieved by RIM_FV_RPP.

From Tables II and III, RIM_FV_RPP performs better than RIM. This shows the necessity of encoding RIM into a more discriminative representation. Compared with RIM_VLAD_RPP, higher APs are achieved by RIM_FV_RPP on airplane and car detection. This demonstrates that the Fisher vector technique is more suitable for the RIM feature. In addition, RIM_FV_RPP performs better than SIFT_FV_RPP on both two data sets. This comparison could demonstrate the higher descriptive ability of RIM. Finally, RIM_FV_RPP performs the best among all the methods on the two data sets, showing its strong discrimination for rotated objects.

(a)



(b)

Fig. 4. Some detection results by RIM_FV_RPP. (a) Airplanes. (b) Cars. True positives are indicated in red and false positives are indicated in cyan, and the undetected objects are indicated by blue bounding box.

The superior performance of our method demonstrates the importance of incorporating rotation invariance in feature design. Fig. 4 shows some airplane and car detection results. Although the extracted proposal centers are not the actual object centers, most objects with arbitrary orientations are detected by RIM_FV_RPP with only a few false positives. This shows the good robustness of RIM_FV_RPP to small translation. As shown in Fig. 4(b), some cars in shadow are successfully detected, demonstrating RIM_FV_RPP's good robustness to the brightness change.

IV. Conclusion

A novel feature is proposed to detect rotated objects in optical remote sensing images. Under the requirement of the rotation invariance, we construct the RIM to incorporate richer spatial information. Instead of directly using RIM to extract features for object detection, we use it as a low-level feature in a feature encoding framework. First, we use RIM to describe overlapping blocks densely partitioned from an image region; then, these RIM features are reduced by PCA and subsequently encoded into Fisher vectors; in the end, a rotation-invariant RPP strategy is used to incorporate spatial information. These well-designed procedures make the proposed feature not only rotation invariant but also with strong discriminative ability. Experiments on airplane/car detection in aerial images show the superiority of our method.

References

[1] W. Zhang, X. Sun, K. Fu, C. Wang, and H. Wang, "Object detection in high-resolution remote sensing images using rotation invariant parts based model," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 1, pp. 74–78, Jan. 2014.

[2] X. Wen, L. Shao, Y. Xue, and W. Fang, "A rapid learning algorithm for vehicle classification," *Inf. Sci.*, vol. 295, pp. 395–406, Feb. 2015.

[3] Z. Zhou, C.-N. Yang, B. Chen, X. Sun, Q. Liu, and Q. M. J. Wu, "Effective and efficient image copy detection with resistance to arbitrary rotation," *IEICE Trans. Inf. Syst.*, vol. E99.D, no. 6, pp. 1531–1540, 2016.

[4] P. Zhou, G. Cheng, Z. Liu, S. Bu, and X. Hu, "Weakly supervised target detection in remote sensing images based on transferred deep features and negative bootstrapping," *Multidimensional Syst. Signal Process.*, vol. 27, no. 4, pp. 925–944, 2016.

[5] X. Chen, S. Xiang, C.-L. Liu, and C.-H. Pan, "Vehicle detection in satellite images by hybrid deep convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 10, pp. 1797–1801, Oct. 2014.

[6] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. CVPR*, 2005, pp. 886–893.

[7] X. Chen, S. Xiang, C.-L. Liu, and C.-H. Pan, "Aircraft detection by deep belief nets," in *Proc. ACPR*, 2013, pp. 54–58.

[8] S. Lazebnik, C. Schmid, and J. Ponce, "A sparse texture representation using local affine regions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1265–1278, Aug. 2005.

[9] Y. Lin, H. He, Z. Yin, and F. Chen, "Rotation-invariant object detection in remote sensing images based on radial-gradient angle," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 4, pp. 746–750, Apr. 2015.

[10] F. Perronnin, J. Sánchez, and T. Mensink, "Improving the Fisher kernel for large-scale image classification," in *Proc. ECCV*, 2010, pp. 143–156.

[11] D. Zhang, J. Han, G. Cheng, Z. Liu, S. Bu, and L. Guo, "Weakly supervised learning for target detection in remote sensing images," *IEEE Geosci. Remote Sens. Lett.*, vol. 12, no. 4, pp. 701–705, Apr. 2015.

[12] M.-M. Cheng, Z. Zhang, W.-Y. Lin, and P. H. S. Torr, "BING: Binarized normed gradients for objectness estimation at 300 fps," in *Proc. CVPR*, 2014, pp. 3286–3293.

[13] G. Heitz and D. Koller, "Learning spatial context: Using stuff to find things," in *European Conference on Computer Vision*. Springer, 2008, pp. 30–43.

[14] G. Cheng and J. Han, "A survey on object detection in optical remote sensing images," *ISPRS J. Photogramm. Remote Sens.*, vol. 117, pp. 11–28, Jul. 2016.

[15] C. L. Zitnick and P. Dollár, "Edge boxes: Locating object proposals from edges," in *Proc. ECCV*, 2014, pp. 391–405.

[16] X. Bai, H. Zhang, and J. Zhou, "VHR object detection based on structural feature extraction and query expansion," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6508–6520, Oct. 2014.

[17] H. Larochelle, D. Erhan, A. C. Courville, J. Bergstra, and Y. Bengio, "An empirical evaluation of deep architectures on problems with many factors of variation," in *Proc. ICML*, 2007, pp. 473–480.

[18] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. ICCV*, 1999, pp. 1150–1157.

[19] R. Arandjelovic and A. Zisserman, "All about VLAD," in *Proc. CVPR*, 2013, pp. 1578–1585.

[20] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," in *Proc. 18th ACM Int. Conf. Multimedia*, 2010, pp. 1469–1472.

[21] R.-E. Fan, K.-W. Chang, C.-J. Hsieh, X.-R. Wang, and C.-J. Lin, "LIBLINEAR: A library for large linear classification," *J. Mach. Learn. Res.*, vol. 9, pp. 1871–1874, Jun. 2008.

[22] A. Vedaldi, M. Blaschko, and A. Zisserman, "Learning equivariant structured output SVM regressors," in *Proc. IEEE ICCV*, Nov. 2011, pp. 959–966.

[23] U. Schmidt and S. Roth, "Learning rotation-aware features: From invariant priors to equivariant descriptors," in *Proc. IEEE CVPR*, Jun. 2012, pp. 2050–2057.

[24] K. Liu *et al.*, "Rotation-invariant HOG descriptors using Fourier analysis in polar and spherical coordinates," *Int. J. Comput. Vis.*, vol. 106, no. 3, pp. 342–364, Feb. 2014.