Online Target Recognition for Time-Sensitive Space Information Networks

Chunlei Huo, *Member, IEEE*, Zhixin Zhou, Kun Ding, and Chunhong Pan

***Abstract*—The key difficulties of online target recognition task for space information networks lie in the contradiction between time-sensitive response requirement and resource constraints(e.g., computation resource, communication resource, and training samples). To deal with the above problems, an effective online target recognizing approach is proposed, which seamlessly integrates fast online information processing task and efficient target-specific high-rate compression task. The proposed approach begins with enhancing the target-background separability by introducing intraclass and interclass couples, the new model adapted for the hospt spot image is then obtained by capturing the relation between the online target data and the massive historical data. The light-weight target-specific information is efficiently transmitted into the ground system, and the whole scene is capable of being recovered while the details of targets are being preserved. Compared with the traditional target recognition methods, the proposed approach is more promising for time-sensitive space information networks. The experiments demonstrate the effectiveness of the proposed approach.**

***Index Terms*—Feature adaption, information processing, information compression, space information networks, target recognition.**

I. Introduction

DETECTING and recognizing targets from images is an important topic for various domains, e.g., remote sensing, pattern recognition, computer vision, and machine learning, etc. Different from traditional target recognition system where images are transmitted to the ground receiving station after being acquired by spaceborne or airborne cameras and the target recognition task is accomplished by the ground processing system, online target recognition system identifies targets from the hot spot image[1] on the imaging platform by embedded systems such as DSP(Digital Signal Processor) and FPGA(Field-Programmable Gate Array), and only the small-size target-related information is sent to the ground receiving station. Online target recognition is more promising for some emergent applications(for instance, searching for MH370) since the response time from image generation to target recognition is reduced significantly.

Space information networks(SIN) are network systems based on various space platforms, such as geostationary satellites, medium and low earth orbit satellites, stratospheric balloons, etc. Increased details provided by high resolution images and rich data acquired by SIN make online target recognition possible. However, online target recognition from high resolution images within SIN is very challenging, and the main challenges boil down to the contradiction between time-sensitive response and resource constraints. To clarify this point, we analyze the challenges from the following two aspects:

1) The low overall separability between target and background, as well as the shortage of training samples, make online target recognition more difficult.

From the perspective of pattern recognition, target recognition is essentially to assign the pixel- or patch-wise features to the label of background or target. In consequence, two fundamental techniques for target recognition are feature descriptor and feature classification. During the past decades, many novel approaches have been proposed in the literature [1]–[6]. High performance target recognition is usually carried out in feature space, and effective feature descriptor can improve the performance significantly [7], [8]. The most widely used feature descriptors include texture features(e.g., HOG [9], SIFT descriptor [10], DAISY [11], LBP [12]), shape features(shape context [13]), superpixel cues [14]), salience features(e.g., objectness [15], BING [16]) or other higher-level feature coding such as bag-of-words [17], sparse coding [18], [19], vector quantization [20], feature learning [21]–[23], etc. The type of features can be determined by template matching(rigid template matching [24], deformable template matching [24], subwindow search [25]), classification(e.g., kNN [26], SVM [27], [28], latent SVM [29], [30], Adaboost [31], CRF [32], neural network [33], [34], Random Forest [35]), knowledge and inference(context knowledge [36], Dempster-Shafer evidential inference [37]), etc. Despite the novelties of above techniques, they are limited in recognizing targets from high resolution images. Specifically, traditional features and classifiers are inadequate for capturing the interclass variability. As illustrated in Fig. 1(a), the feature distance between points A and B is even higher than that between

Manuscript received June 29, 2016; revised October 23, 2016; accepted January 10, 2017. Date of publication January 18, 2017; date of current version May 8, 2017. This work was supported in part by the Natural Science Foundation of China under Grant 91438105, Grant 61375024, Grant 61302170, and Grant 91338202. The guest editor coordinating the review of this manuscript and approving it for publication was Prof. Eric Miller.

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

Z. Zhou is with Beijing Institute of Remote Sensing, Beijing 100191, China (e-mail: zhixin.zhou@mail.ia.ac.cn).

K. Ding is with National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China (e-mail: kding@nlpr.ia.ac.cn).

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

Digital Object Identifier 10.1109/TCI.2017.2655448

[1]The hot spot image is the image acquired over the hot spot area that attracts extensive interests. For target recognition application, the hot spot image means the image that contains the targets to be recognized.

Fig. 1. Illustration of difficulties in online target recognition. (a): Historical image. The feature distance between points A and B is even higher than that between points B and C, i.e., low interclass variability and high intraclass difference. (b): Hot spot image. There exist complex spectral differences between historical images and the hot spot image, and the model learned from the historical images cannot be directly applied to the hot spot image.

points B and C, and the performance will be impacted by this ambiguity. Furthermore, there exist complex spectral differences between the historical images (Fig. 1(a)) and the hot spot image (Fig. 1(b)), and the lack of training samples from the hot spot image makes it difficult to refine the model trained from the historical images.

2) The contradiction between time-sensitive response requirement and resource constraints makes the traditional remote sensing information system(RSIS) inappropriate to SIN.

Generally, the traditional RSIS is composed of three modules: imaging module, data communication module and ground system. For SIN, one main drawback of RSIS is the lack of cooperation between the above modules. In detail, high resolution images are too large in sizes to be transmitted by the bandwidth-limited communication network, and the massive amounts of data are prohibitive for the ground system. The above contradictions require online target recognition and light-weight transmission. On one hand, the above changes require online processing algorithms achieve the satisfying performance with limited computation resource and training data but with higher efficiency. On the other hand, new compression techniques should be developed to reduce the communication pressure. Although many novel compressing techniques (e.g., DCT compression, wavelet compression, compressed sensing, etc.) have been proposed, the compression rate and the computation complexity are inadequate for SIN. Moreover, they ignore the differences between targets and background.

In short, there are significant differences between the traditional RSIS and the newly developed SIN, and it is urgent for the researchers to develop innovative approaches to tackle the above problems. To this aim, a novel online target recognition approach is proposed. Compared with the traditional techniques, the novelties of the proposed approach lie in the following two aspects: (a) fast online target recognition by learning the relationship between interclass and intraclass features and that between the hot spot data and historical data, (b) light-weight information compression by target-specific non-uniform sampling and target-preserved image recovery. In this paper, light-weight information compression means not only high compression rate, but also simple and fast compression procedure.

II. The Proposed Approach

This paper is aimed at addressing the target recognition and information compression problems in the context of SIN. The rationale of the the proposed approach is to (1) learn a discriminative metric matrix by mining relationship between intraclass and interclass features, as well as the relationship between historical images and hot spot images. (2) extract target-related small-size features by which the hot spot image can be recovered while the details of targets can be preserved. As illustrated by Fig. 2, the proposed approach consists of four steps: offline relationship learning, online target recognition, online information compression and offline information recovery. Below, we elaborate the proposed technique step by step.

A. Offline Relationship Learning

As stated before, the first difficulty in target recognition is the low variability between targets and the clutter background, as well as the low separation between targets of different types. In this paper, the problem is addressed by capturing the interclass difference and intraclass similarity, which are defined by the following two concepts:

Def. 1 Intraclass couple: The sample pair $(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is defined as an intraclass couple if $\boldsymbol{x}_j$ is one of the nearest neighbors of $\boldsymbol{x}_i$ and shares the same class label with $\boldsymbol{x}_i$, i.e., $\boldsymbol{x}_j \in \mathcal{N}(\boldsymbol{x}_i)$ and $y_i = y_j$, y_i and y_j are the labels of $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, respectively. $\mathcal{N}(\boldsymbol{x}_i)$ means the neighborhood of $\boldsymbol{x}_i$. $\boldsymbol{x}_j$ is called the **target neighbor** of $\boldsymbol{x}_i$.

Def. 2 Interclass couple: The sample pair $(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is called an interclass couple if $\boldsymbol{x}_j$ is one of the nearest neighbors of the sample $\boldsymbol{x}_i$ but has different class label with $\boldsymbol{x}_i$, i.e., $\boldsymbol{x}_i \in \mathcal{N}(\boldsymbol{x}_j)$ and $y_i \neq y_j$. $\boldsymbol{x}_j$ is called the **impostor** of $\boldsymbol{x}_i$.

Intraclass couples and Interclass couples are illustrated in Fig. 3. Given a couple of training samples $(\boldsymbol{x}_i, y_i)$ and $(\boldsymbol{x}_j, y_j)$, the overall variability is aimed to be improved by a metric matrix $\boldsymbol{M}$ as follows:

$$d_{\boldsymbol{M}}(\boldsymbol{x}_i, \boldsymbol{x}_j) = (\boldsymbol{x}_i - \boldsymbol{x}_j)^T \boldsymbol{M} (\boldsymbol{x}_i - \boldsymbol{x}_j). \tag{1}$$

Where $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are feature vectors, and y_i and y_j are the corresponding labels. Generally, $d_{\boldsymbol{M}}(\boldsymbol{x}_i, \boldsymbol{x}_j) \leq \tau_1$ is expected to hold if $(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is an intraclass couple, and $d_{\boldsymbol{M}}(\boldsymbol{x}_i, \boldsymbol{x}_j) > \tau_2$ if $(\boldsymbol{x}_i, \boldsymbol{x}_j)$ is an interclass couple, where τ_1 is a relatively small value and τ_2 is a sufficiently large one. In other words, the distance between features should not be computed directly based on the feature difference, but learned and tuned adaptively driven by the relationship hidden in the training samples. The term "relationship" means the relation between targets and background and the relation between the historical and hot spot images. In this paper, the metric matrix $\boldsymbol{M}$ is learned simultaneously with the classifier, i.e.,

$$\begin{aligned} &\min_{\boldsymbol{M}, \xi_l} \frac{1}{2} \|\boldsymbol{M} - \boldsymbol{I}\|_F^2 + C \sum_l \xi_l \\ &s.t.\ \ h_l((\boldsymbol{x}_{l,1} - \boldsymbol{x}_{l,2})^T \boldsymbol{M} (\boldsymbol{x}_{l,1} - \boldsymbol{x}_{l,2})) \geq 1 - \xi_l, \\ &\xi_l \geq 0, \forall l. \end{aligned} \tag{2}$$

Where $\|\cdot\|_F$ denotes the Frobenius norm, and $\boldsymbol{I}$ is the identity matrix. ξ_l denotes a slack variable, and C is the

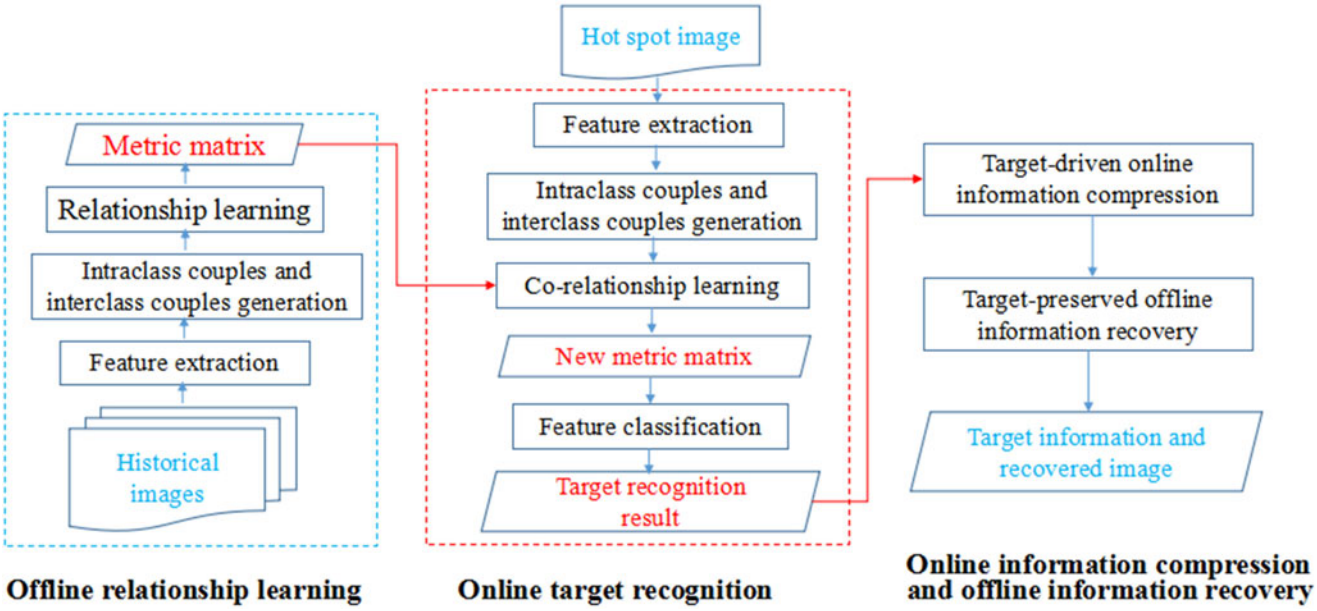


Fig. 2. Diagram of proposed approach. The proposed approach consists of four steps: offline relationship learning, online target recognition, online information compression and offline information recovery. Offline relationship learning learns metric matrix from historical images, and online target recognition learns new metric matrix for the hot spot image driven by few new training features and the relationship between metric matrices. Online information compression extracts target-related small-size features, and offline information recovery reproduces the target recognition result and the hot spot image.

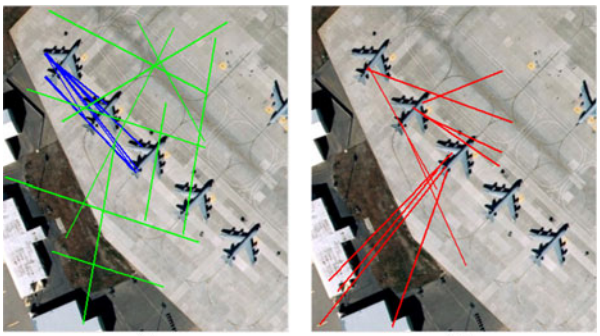


Fig. 3. Illustration intraclass couples and Interclass couples. (a) : Intraclass couples. (b): Interclass couples.

regularization factor. $h_l = -1$ if $(\boldsymbol{x}_{l,1}, \boldsymbol{x}_{l,2})$ is an intraclass couple, and $h_l = 1$ if it is an interclass couple. Intraclass couples can be generated by combining training features with same labels, and interclass couples generated by combining training features with different labels. The optimization procedure can be referred to Appendix A.

Once the metric matrix $\boldsymbol{M}$ is obtained, the label for each test feature $\boldsymbol{x}$ is determined by

$$\ell(\boldsymbol{x}) = sgn\left(\sum_{i=1}^{k} \alpha_i y_i \boldsymbol{d}_{\boldsymbol{M}}(\boldsymbol{x_i}, \boldsymbol{x})\right). \tag{3}$$

Where $\boldsymbol{x_i}$ denotes the ith nearest neighbor of $\boldsymbol{x}$ within training samples, and $\boldsymbol{\alpha_i}$ is the weight of $\boldsymbol{x_i}$ and it can be computed by solving Eq. (19) in Appendix. k is the number of nearest neighbors.

For the immense historical data, the metric matrix can be trained offline with the help of plentiful training samples, where a training sample is consisted of a pixel-wise feature and the corresponding label. In this paper, DAISY feature is used due to its powerful representative ability and high efficiency. Specifically, DAISY feature is a dense and representative feature descriptor, and DAISY feature is extracted at each pixel. For an image of 1000*1000 pixels, the computation time of DAISY is about only 10 seconds. Other descriptors can be utilized without problem, but the comparisons of different features are beyond the scope of this paper.

B. Online Target Recognition

Despite powerful discriminative abilities of $\boldsymbol{M}$, there exist significant differences between the hot spot image and the historical images, and the available training samples for the hot spot image are scarce. For this reason, we propose utilizing co-relationship learning to establish the relation between the hot spot image and the historical images with the help of few training samples available from the hot spot image.

Suppose we are given n labeled training samples $\{(\mathbf{x}_j, y_j)|j = 1, \cdots, n\}$, where $n \ll N$, and N denotes the total number of features extracted from the hot spot image, co-relationship learning is to learn the new metric matrix $\boldsymbol{Q}$ and the relationship between metric matrices:

$$\min_{\boldsymbol{Q},\Omega} \frac{1}{2}\|\boldsymbol{Q} - \boldsymbol{I}\|_F^2 + \lambda_1 tr(\widetilde{v}\Omega^{-1}\widetilde{v}^T) + \lambda_2 \sum_{j<k}[h_{j,k}((\mathbf{x}_j - \mathbf{x}_k)^T \boldsymbol{Q}(\mathbf{x}_j - \mathbf{x}_k)) - 1]$$

$$s.t.\ \boldsymbol{Q} \geq \mathbf{0}$$
$$\widetilde{v} = (vec(\boldsymbol{M}), vec(\boldsymbol{Q}))$$
$$\Omega = \begin{pmatrix} 1-\omega_1 & \omega_2 \\ \omega_2 & \omega_1 \end{pmatrix}$$
$$\omega_1(1-\omega_1) \geq \omega_2^2 \tag{4}$$

Where $vec(\cdot)$ denotes the operator which converts a matrix into a vector in a column-wise manner. Ω is a covariance matrix which describes the relationship between metric matrices. ω_2 denotes the covariance between $\boldsymbol{M}$ and $\boldsymbol{Q}$, and ω_1 the variance of $\boldsymbol{Q}$. λ_1 and λ_2 are the regularization factors.

Similar to [38], the above problem can be solved by alternative updating between $\boldsymbol{Q}$ and ω_i(i = 1, 2) . Specifically, we first optimize the objective function with respect to $\boldsymbol{Q}$ when ω_1 and ω_2 are fixed, and then optimize the objective function with respect to ω_1 and ω_2 when $\boldsymbol{Q}$ is fixed. This procedure is repeated until convergence. Noting that in Eqs. 2 and 4, $\|\boldsymbol{M} - \boldsymbol{I}\|_F^2$ and $\|\boldsymbol{Q} - \boldsymbol{I}\|_F^2$ are to be optimized instead of $\|\boldsymbol{M}\|_F^2$ and $\|\boldsymbol{Q}\|_F^2$, which is different from [38]. The roles of $\|\boldsymbol{M} - \boldsymbol{I}\|_F^2$ and $\|\boldsymbol{Q} - \boldsymbol{I}\|_F^2$ are to prevent the transformed features from being distorted too much and to enhance the stability of the solution(Distorted transformation means a high generalization error.).

After learning the reliable metric matrix $\boldsymbol{Q}$ for online target recognition task, we can make prediction for the hot spot image. Given a test feature $\boldsymbol{x}_t$, we first calculate the distances between $\boldsymbol{x}_t$ and online training samples based on the learned metric $\boldsymbol{Q}$ and then use the $k-$nearest neighbor classifier to classify $\boldsymbol{x}_t$. The role of k is to remove the isolated pixels and make the target region smooth, but too large k will produce false alarms. By experiments, we found best performances can be achieved at $k = 3$.

C. Target-Driven Information Compression

Due to the bandwidth limitation of SIN, it is difficult to promptly transmit the large-size high resolution image to the ground system. The positions and types of targets recognized above are small in size, and they can be transmitted to the ground system. However, it is difficult to validate targets without the adjacent context information. For this reason, besides small-size target information, the light-weight necessary information must be transmitted to the ground station for recovering the hot spot image.

In this paper, the light-weight necessary information is generated by nonuniform sampling adaptively driven by the target recognition result, i.e., the target areas are sampled in a denser rate(e.g., 10%), and the background areas in a sparser rate (e.g., 3%). In detail, let $\boldsymbol{u}_i \in R^P (1 \leq i \leq n)$ be the spectral response vector extracted from the hot spot image at the ith pixel, where P and n denote the number of channels and pixels, respectively. The sampling procedure is expressed as

$$\boldsymbol{v}_i = \boldsymbol{\Sigma}\boldsymbol{u}_i \tag{5}$$

Where the element of $\boldsymbol{\Sigma}$ is a binary alphabet to denote whether the corresponding pixel is chosen. For example, $\boldsymbol{\Sigma}_i = 1$ means the ith pixel is chosen, and $\boldsymbol{\Sigma}_i = 0$ means the ith pixel not chosen. To further reduce the communication pressure, the above information is encoded in a sparse format, $(k, \boldsymbol{u}_k)$, k is the position index of the chosen pixel, and $\boldsymbol{u}_k$ the spectral response vector. In fact, only contour positions of target areas and the above compressed information must be transmitted, since the details of targets and the background can be recovered after they are send to the ground processing system. It is worth noting that target-driven information compression is fast since it is essentially the sampling operator.

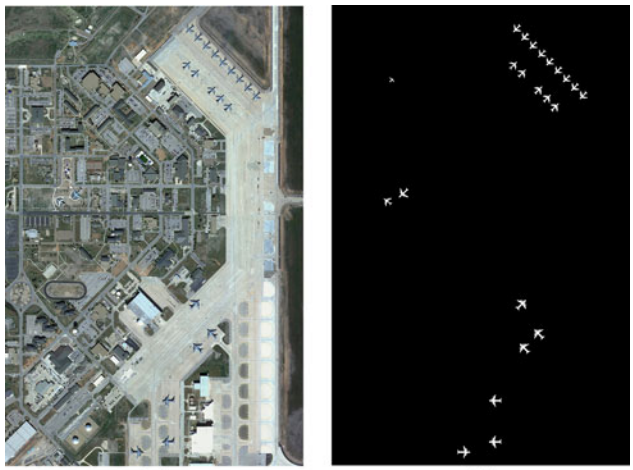


Fig. 4. Historical image and reference ground truth. (a) Historical image(2732×1773). (b): Reference ground truth.

D. Target-Preserved Information Recovery

In this paper, the hot spot image is recovered by applying dictionary learning and compressive sensing techniques on the light-weight context information. In detail, suppose the original hot spot image $X \in R^{N \times P}$ is expressed by the following dictionary learning problem:

$$\boldsymbol{u}_i = \boldsymbol{D}\boldsymbol{w}_i + \boldsymbol{\epsilon}_i \tag{6}$$

Where $\boldsymbol{w}_i \in R^K$ is the sparse coefficient of X at the ith pixel or patch, and the columns of matrix $\boldsymbol{D} \in R^{P \times K}$ represents the K components of a dictionary with which $\boldsymbol{u}_i$ is expanded. $\boldsymbol{\epsilon}_i$ is the measurement noise and reconstruction error. From Eqs. (5) and (6), it can be observed that $\boldsymbol{v}_i = \boldsymbol{\Sigma}(\boldsymbol{D}\boldsymbol{w}_i + \boldsymbol{\epsilon}_i)$, with $\boldsymbol{\Phi} = \boldsymbol{\Sigma}\boldsymbol{D}$ and $\boldsymbol{t}_i = \boldsymbol{\Sigma}\boldsymbol{\epsilon}_i$.

Motivated by [39], $\boldsymbol{u}_i$ is recovered from $\boldsymbol{v}_i$, $\boldsymbol{D}$ and $\{\boldsymbol{w}_i\}_{i=1,\cdots,N}$ by the following models:

$$\boldsymbol{v}_i = \Phi\boldsymbol{u}_i + \boldsymbol{t}_i \tag{7}$$
$$\boldsymbol{w}_i = \boldsymbol{z}_i \odot \boldsymbol{s}_i \tag{8}$$
$$\boldsymbol{d}_k \sim \mathcal{N}(0, P^{-1}I_P) \tag{9}$$
$$\boldsymbol{s}_i \sim \mathcal{N}(0, \gamma_s^{-1}I_K) \tag{10}$$
$$\boldsymbol{\epsilon}_i \sim \mathcal{N}(0, \gamma_\epsilon^{-1}I_P) \tag{11}$$

Where $\boldsymbol{d}_k$ represents the ith component (atom) of $\boldsymbol{D}$, $\odot$ represents the element-wise or Hadamard vector

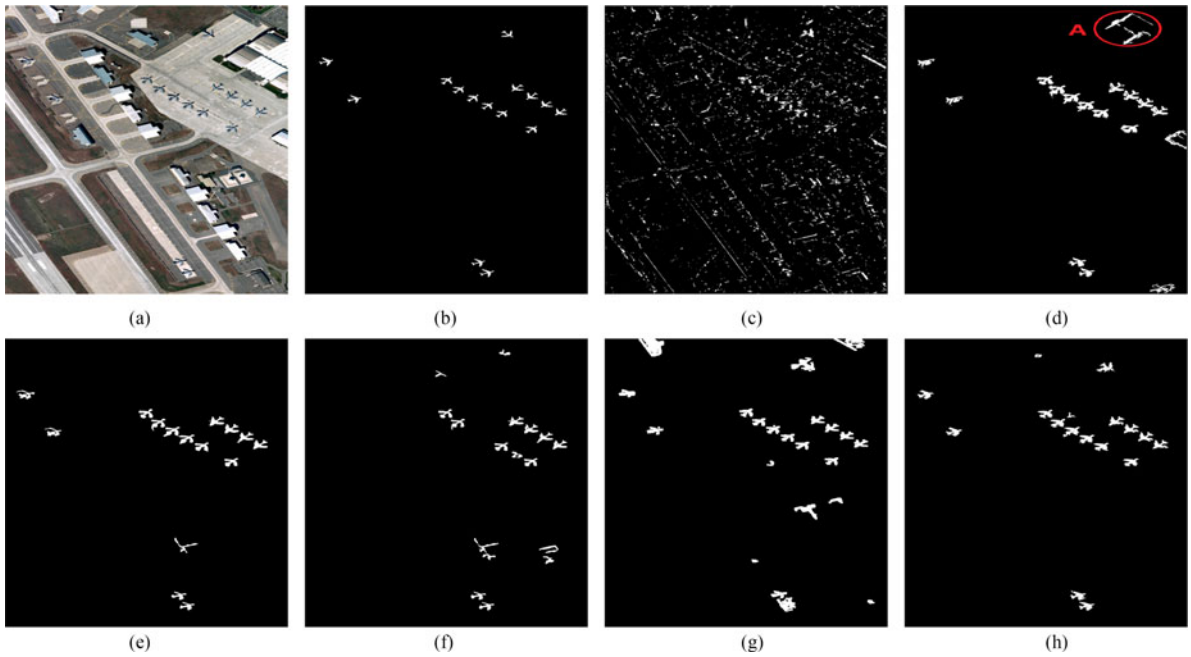


Fig. 5. Results comparison on dataset 1. (a): Hot-spot image1(size: 1975 × 1106). (b): Reference ground truth, (c): SVM, (d): DML, (e): MMDT, (f): AdaSVM, (g):SalRDS, (h):RLAS.

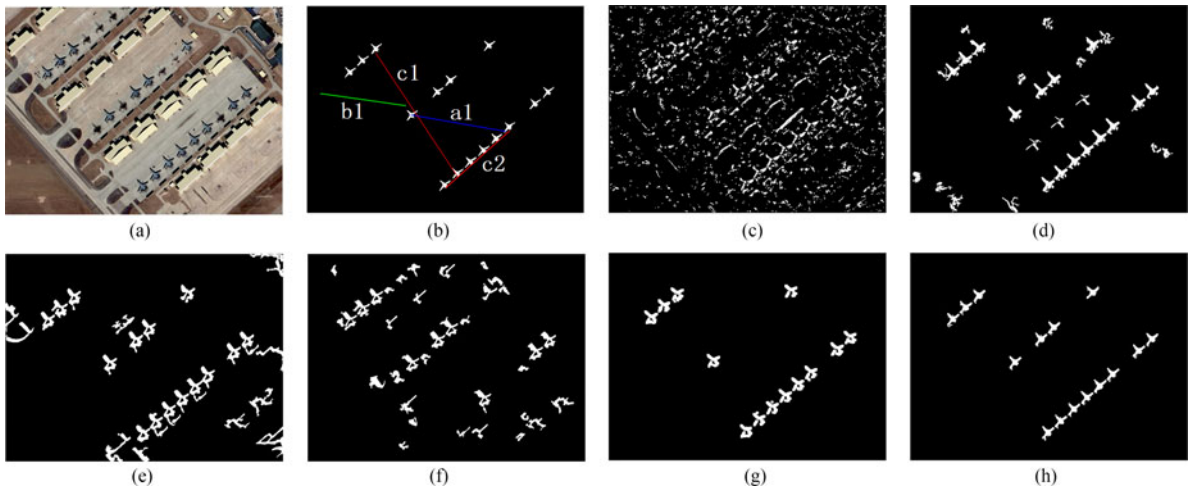


Fig. 6. Results comparison on dataset 2. (a): Hot-spot image1(size: 1135 × 850). (b): Reference ground truth, (c): SVM, (d): DML, (e): MMDT, (f): AdaSVM, (g):SalRDS, (h):RLAS. In Fig. 6(b), a1 and b1 denote intraclass couples, and c1 and c2 are interclass couples.

product, $I_P(I_K)$represents a $P \times P(K \times K)$identity matrix, $\{\boldsymbol{z}_i\}_{i=1,\cdots,n}$ are drawn as in (5). Conjugate hyper priors $\gamma_s \sim \Gamma(c,d)$ and $\gamma_\epsilon \sim \Gamma(e,f)$. $\boldsymbol{z}_i \in \{0,1\}^K$ is a binary vector defining which columns of D are used for representing $\boldsymbol{u}_i$. Hence, for data $\{\boldsymbol{u}_i\}_{i=1,\cdots,n}$, there is an associated set of latent binary vectors $\{\boldsymbol{z}_i\}_{i=1,\cdots,n}$, and the Beta-Bernoulli process provides a convenient prior for these vectors. The details can be referred to [39].

III. Experiments

A. Experiments Description

To validate the effectiveness of the proposed approach, we simulated online target recognition task in SIN enviroment and conducted many experiments on different datasets. Owing to space limitations, only the results for three datasets are illustrated in this paper. In this paper, each dataset consists of a high resolution image acquired by QuickBird2. One image with accurate labels (Fig. 4) is considered as the historical data, and other three images (Figs. 5–7) of different sites are used as the hot spot images, respectively.

The main advantages of the proposed approach lie in online target recognition and light-weight compression. For this reason, two groups of experiments are designed. The first group is to validate the effectiveness of relationship learning for online target recognition. To this aim, the proposed approach is compared to the following five related approaches:

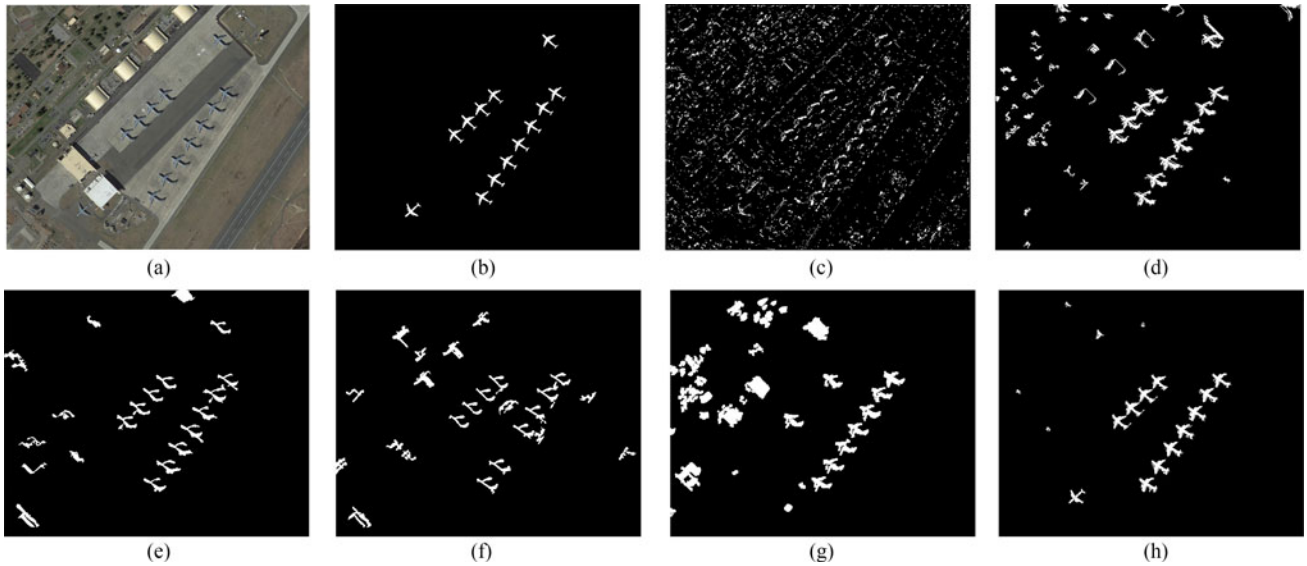


Fig. 7. Results comparison on dataset 3. (a): Hot-spot image1(size: 1454 × 1802). (b): Reference ground truth, (c): SVM, (d): DML, (e): MMDT, (f): AdaSVM, (g):SalRDS, (h):RLAS.

1) *SVM:* SVM is used to train the model on the massive historical data, and the model is directly used for classifying the features extracted from the hot spot image.

2) *DML(Direct Metric Learning on hot spot image):* Limited number of training samples are used to train the model for the hot spot image based on Eq. (2).

3) *MMDT(Max Margin Domain Transform):*[40]. It is aimed at optimizing the linear transformation that maps features from the hot spot image to the historical image, and the transformation and classifier parameters are jointly solved by the variant of SVM.

4) *AdaSVM(Adaptive SVM):* [41]. It transforms existing classifier into an effective classifier by learning a delta function between the original and adapted classifier using an objective function similar to SVM. Noting that AdaSVM was performed on the original features instead of interclass and intraclass couples.

5) *SalRDS(Salient Region Detection and Segmentation) [42]:* It calculates salient map using low level features of color and luminance, and segments salient objects by K-means.

For convenience, the proposed target recognition approach is abbreviated to **RLAS**(Relationship Learning and Adaptive Sampling). For SVM and DML, one regularization parameter needs tuning. For MMDT, AdaSVM and RLAS, two regularization parameters need tuning. The above parameters are tuned on the training set by 5-fold cross-validation. RBF kernel is used for SVM, MMDT and AdaSVM. The target recognition performances of different approaches are evaluated by FNR(False Negative Rate), FPR(False Positive Rate), OA(Overall Accuracy), Kappa coefficient and CPU time on online procedures. In this paper, we use LIBSVM [43] for SVM and use the source codes provided by the authors for MMDT, AdaSVM and SalRDS. DML and RLAS are implemented in Matlab. The algorithms were tested on a desktop computer (Intel Core i7 920 Quad-Core CPU, 2.67 GHz, 8-GB DDR RAM).

TABLE I
RECOGNITION PERFORMANCE COMPARISON

Dataset	Approach	FNR(%)	FPR(%)	OA(%)	Kappa	CPU(s)
1	SVM	67.63	3.64	81.99	0.08	131
	DML	11.76	5.07	90.87	0.32	176
	MMDT	9.17	2.84	91.11	0.54	361
	AdaSVM	35.45	2.83	91.97	0.42	186
	SalRDS	16.90	2.91	92.00	0.32	13
	RLAS	2.27	1.72	93.28	0.62	162
2	SVM	57.45	11.71	83.77	0.11	60
	DML	5.25	7.09	90.94	0.39	143
	MMDT	15.43	5.52	89.38	0.22	172
	AdaSVM	45.04	6.06	86.53	0.19	151
	SalRDS	21.43	3.62	91.08	0.41	11
	RLAS	2.11	1.72	94.28	0.73	121
3	SVM	74.84	7.78	80.39	0.10	167
	DML	13.18	5.08	88.80	0.37	184
	MMDT	36.43	4.17	91.44	0.35	459
	AdaSVM	44.62	3.34	90.05	0.25	210
	SalRDS	36.89	6.43	89.13	0.17	14
	RLAS	6.53	1.81	93.10	0.70	176

The second group of experiments are to demonstrate the efficiency and effectiveness of online information compression technique. To this aim, the proposed approach is compared to other prominent approaches quantitatively and qualitatively: PCA [44], SVD [45], NMF [46]. The compression performances are measured by compression time, compression rate(the ratio of the compressed size to the original size) and the recovery fidelity. In this paper, the recovery fidelity is evaluated by LCor, the local correlation within the target areas.

B. Experiments Analysis

The results of different target recognition approaches are shown in Figs. 5– 7, and the performances are listed in Table I. From Table I, it can be informed that SVM performs worst.

Fig. 8. Recovery results comparison. First column: the proposed approach, Second column: SVD, Third column: NMF, Fourth column: PCA.

For instance, on dataset 1, it obtains highest FNR(67.73%) and lowest OA(81.99%) and Kappa(0.08). This can be verified by checking Fig. 5(c), where many pixels are misclassified as targets, and detected targets are hard to be discriminated from the background. The underlying reason lies in the fact that the significant differences between the historical data and hot spot data were ignored, and the classification performance cannot be guaranteed even if the model is trained accurately from the massive historical data. Furthermore, SVM is even worse than SalRDS, a salience-based approach. However, once the relationship is considered, the performance is improved significantly even if the relationship between the historical data and hot spot data is established based on the limited number of training samples extracted from the hot spot image. For instance, on dataset1, Kappa is improved from 0.08 by SVM to 0.54 by MMDT and to 0.42 by AdaSVM. Similarly, on dataset3, Kappa is improved from 0.10 by SVM to 0.35 by MMDT and to 0.25 by AdaSVM. The above performance improvements demonstrate the importance and effectiveness of feature transformation(MMDT) or classifier adaption(AdaSVM). By comparing the results achieved by MMDT and AdaSVM, we can observe that FNR and FPR are much smaller than that by SVM. However, as stated in Section I, the variabilities between targets and the clutter background are low, and they are difficult to be separated even with the help of feature transformation and classifier adaption, if the relationship between interclass and intraclass features is neglected. For this reason, the performance obtained by DML is even higher than MMDT and AdaSVM. For example, on dataset 2, Kappa is improved from 0.22 by MMDT and 0.19 by AdaSVM to 0.39 by DML.

Noting that on dataset 1, DML is inferior to MMDT and AdaSVM, the main reasons lie in the following two facts:

1) The small training samples from the hot spot image is inadequate for training an accurate model, and the generated intraclass and interclass training samples fail to capture the subtle difference between targets and background. For instance, the shape of the region A in Fig. 5(d) is similar to an airplane.

2) The similarities between the historical image and dataset 1 are higher than that between the historical image and dataset 2 and dataset 3. In consequence, the performance improvements taken by MMDT and AdaSVM on dataset 1 are higher than that on the latter two datasets.

The above two facts illustrate the importance of interclass and interclass couples in improving the overall variabilities between targets and background. Moreover, its ability is enhanced significantly by establishing the relations between the hot spot

TABLE II
DISTANCE COMPARISON BEFORE AND AFTER RELATIONSHIP LEARNING

interclass(intraclass) couple	distance before relationship learning	distance after relationship learning
a1	0.57	0.11
b1	1.38	0.32
c1	0.19	1.23
c2	0.20	0.98

image and the historical image. For example, Kappas obtained by RLAS are 0.62, 0.73 and 0.70, respectively, which are much higher than other approaches. The improved performance is contributed to the combination on relationship learning and transfer learning. Noting that in transfer learning, we learn the relationship between metric matrices, which is different from MMDT and AdaSVM. Despite limited training samples from the hot spot image used for training the new metric matrix, by taking advantages of the constraints between two metric matrices, accurate relationship between interclass and intraclass couples is transferred to the new metric matrix, which otherwise will be suffered from the shortage of adequate training samples. The other advantage of the proposed approach lies in the edge-preserved ability near the target boundaries. In detail, interclass and intraclass couples capture the difference between targets and background as well as the similarities within targets, and the learned metric matrix is helpful for discriminating targets from the neighboring pixels within the background. As a result, the shapes and edges of the targets are being well preserved.

To understand how relationship learning help improve the overall variabilities, some intraclass couples and interclass couples are shown in Fig. 6(b), and the corresponding distances before and after relationship learning are listed in Table II, where the distances before relationship learning are measured by Euclidean distance between feature couple. From Table II, it can be found that before relationship learning, the distances between interclass couples are smaller than that between intraclass couples. For instance, the distance of intraclass couple b1 is 1.38, and the distance of interclass couple c1 is 0.19. In this case, targets and background are difficult to be separated. However, after relationship learning, the distances of interclass couples are much larger than the distances of intraclass couples. Specifically, the distance of interclass couple c1 and c2 are increased to 1.23 and 0.98, which are larger than the distance of intraclass couples a1 and b1, 0.11 and 0.32. For SVM-like methods, the Gaussian kernel leads to an ordering function that is equivalent to using the Euclidean metric. In consequence, the features at above positions are misclassified by SVM, MMDT and AdaSVM. The distance changes before and after relationship learning demonstrate the importance of relationship learning for improving intraclass similarities and interclass differences.

Regarding online target recognition time, SalRDS is superior to other methods since it is based on simple spectral-based salience features and no advanced training procedures are involved. For this reason, it is limited in reliably recognizing the targets from the clutter background and considering the user-specific interests, this can be validated from Fig. 7(g). SVM is fast since the model was trained offline, and it only extracts and classifies features online. In contrast, DML trains the model online, and it is slower than SVM. For instance, for dataset 2, the computation time of DML is 143s, and SVM is 60s. However, with the image size increase, the time differences are reduced since the improved overall separability taken by DML is helpful for the convergence. For example, the time difference between DML and SVM on dataset 3 is only 17s. MMDT and AdaSVM outperforms SVM in recognition accuracy since they update the model trained offline driven by features extracted from the hot spot image, however, they are slower than SVM and even DML. The underlying reason is the ignorance of feature relationship, i.e., low separability will cost more time to complete the iteration procedure in training step. For this reason, RLAS achieved the best balance between recognition accuracy and computation time.

TABLE III
COMPRESSION PERFORMANCE COMPARISON

Dataset	Approach	Time(s)	Comp. Rate	LCor
	RLAS	0.45	0.05	0.80
	PCA	29.7	0.05	0.58
1	SVD	428.4	0.13	0.59
	NMF	467.4	0.05	0.62
	RLAS	0.19	0.09	0.93
	PCA	9.3	0.07	0.70
2	SVD	228.9	0.21	0.66
	NMF	218.1	0.06	0.71
	RLAS	0.57	0.08	0.92
	PCA	24.9	0.04	0.86
3	SVD	495.3	0.12	0.83
	NMF	305.1	0.04	0.84

The images recovered by different approaches are shown in Fig. 8, and the compression performances are listed in Table III. From Table III, it can be inferred that the proposed approach achieved the best performance with respect to computation time, compression rate and recovery fidelity. Taking dataset 1 for an example, RLAS, the proposed approach achieved the highest compression rate (0.05) and LCor(0.80) with the least computation time (0.45s), while other three approaches are inferior to RLAS with respect to computation time or target fidelity. For instance, for PCA, the compression time is 29.7s, and LCor 0.58. Similarly, the compression time of NMF is 467.4s, and LCor 0.62. Noting that RLAS is very fast in the online information compression procedure, however, it is time-consuming in the latter information recovery step. Considering the fact that compression time is very important for online information compression and light-weight transmission, and the information recovery task is performed by the ground processing system, RLAS is still the most promising for the time-sensitive SIN.

IV. CONCLUSION

The challenges of online target recognition for SIN lie in the contradictions between time-sensitive requirements and the lacks of the computation, communication and prior resources. This paper is aimed at solving the above difficulties

by relationship learning and target-specific information compression. Compared with the related work, the contributions of the proposed approach are two-folds: relationship learning is helpful to improve the overall separability between targets and background and to capture the relationship between the historical data and the hot spot data, and target-specific adaptive compression is promising for fast information compression and prompt light-weight information transmission. Despite the novelties of the proposed approach, more efforts should be directed to develop advanced techniques for SIN. Our future work will focus on more complex target recognition task for resource-constrained SIN.

V. APPENDIX

A. Appendix A

To solve the above problem (2), the Lagrangian version is derived as follows:

$$L(\boldsymbol{M},\xi,\alpha,\beta)=\frac{1}{2}\|\boldsymbol{M}-\boldsymbol{I}\|_F^2+C\sum_l\xi_l$$
$$-\sum_l\alpha_l[h_l((\boldsymbol{x}_{l,1}-\boldsymbol{x}_{l,2})^T\boldsymbol{M}(\boldsymbol{x}_{l,1}-\boldsymbol{x}_{l,2}))-1+\xi_l]$$
$$-\sum_l\beta_l\xi_l, \tag{12}$$

Where α and β are the Lagrange multipliers that satisfy $\alpha\geq 0$ and $\beta\geq 0,\forall l$. To convert the original problem to its dual version, we set the derivative of the Lagrangian version with respect to $\boldsymbol{M}$ and ξ to be 0 :

$$\frac{\partial L(\boldsymbol{M},\xi,\alpha,\beta)}{\partial \boldsymbol{M}}=\boldsymbol{0} \tag{13}$$
$$\Rightarrow(\boldsymbol{M}-\boldsymbol{I})-\sum_l\alpha_l h_l(\boldsymbol{x}_{l,1}-\boldsymbol{x}_{l,2})(\boldsymbol{x}_{l,1}-\boldsymbol{x}_{l,2})^T=\boldsymbol{0}. \tag{14}$$
$$\frac{\partial L(\boldsymbol{M},\xi,\alpha,\beta)}{\partial \xi_l}=0 \tag{15}$$
$$\Rightarrow C-\alpha_l-\beta_l=0 \tag{16}$$
$$\Rightarrow 0<\alpha_l<C,\forall l. \tag{17}$$

Eq. (13) implies that the relationship between $\boldsymbol{M}$ and α can be represented as follows:

$$\boldsymbol{M}=\boldsymbol{I}+\sum_l\alpha_l h_l(\boldsymbol{x}_{l,1}-\boldsymbol{x}_{l,2})(\boldsymbol{x}_{l,1}-\boldsymbol{x}_{l,2})^T. \tag{18}$$

Substituting (13)–(15) back into $L(\boldsymbol{M},\xi,\alpha,\beta)$, we obtain the Lagrange dual problem as follows:

$$\max_\alpha -\frac{1}{2}\sum_{i,j}\alpha_i\alpha_j h_i h_j\boldsymbol{K}_{\boldsymbol{D}}(\boldsymbol{z}_i,\boldsymbol{z}_j)+\sum_i\alpha_i(1-h_i\boldsymbol{z}_i^T\boldsymbol{z}_i) \tag{19}$$
$$s.t.\ 0\leq\alpha_l\leq C,\forall l,$$

where $\boldsymbol{K}_{\boldsymbol{D}}(\boldsymbol{z}_i,\boldsymbol{z}_j)=[\boldsymbol{z}_i^T\boldsymbol{z}_j]^2$, and $\boldsymbol{z}_i=\boldsymbol{x}_{i,1}-\boldsymbol{x}_{i,2}$. The above problem is a standard quadratic program, and it can be solved by a variety of approaches, such as the interior point method, active set method, etc. [47]. Considering the similarity between Eq.(16) and the Lagrange dual problem of Eq.(2), the above problem is solved using LibSVM [43].

REFERENCES

[1] A. Andreopoulos and J. K. Tsotsos, "50 years of object recognition: Directions forward," *Comput. Vis. Image Understanding*, vol. 117, no. 8, pp. 827–891, 2013.

[2] X. Zhang, Y. Yang, Z. Han, H. Wang, and C. Gao, "Object class detection: A survey," *ACM Comput. Surveys*, vol. 46, no. 1, pp. 10:1–10:53, 2013.

[3] R. Verschae and J. Ruiz del Solar, "Object detection: Current and future directions," *Frontiers Robot*. AI, vol. 2, no. 29, pp. 1–7, 2015.

[4] B. Bhanu, "Automatic target recognition: State of the art survey," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, no. 4, pp. 364–379, Jul. 1986.

[5] 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, 2016.

[6] A. Borji, M. Cheng, H. Jiang, and J. Li, "Salient object detection: A survey," *Eprint Arxiv*, vol. 16, no. 7, pp. 3118–3143, 2014.

[7] Y. Li, S. Wang, Q. Tian, and X. Ding, "Feature representation for statistical-learning-based object detection: A review," *Pattern Recognit.*, vol. 48, pp. 3542–3559, 2015.

[8] Y. Huang, Z. Wu, L. Wang, and T. Tan, "Feature coding in image classification: A comprehensive study," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 3, pp. 493–506, Mar. 2014.

[9] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2005, pp. 886–893.

[10] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[11] E. Tola, V. Lepetit, and P. Fua, "A fast local descriptor for dense matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2008, pp. 1–8.

[12] T. Ojala, M. Pietikainen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognit.*, vol. 29, no. 1, pp. 51–59, 1996.

[13] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, Apr. 2002.

[14] Y. Zhong, R. Gao, and L. Zhang, "Multiscale and multifeature normalized cut segmentation for high spatial resolution remote sensing imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 6061–6075, Oct. 2016.

[15] B. Alexe, T. Deselaers, and V. Ferrari, "Measuring the objectness of image windows," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 11, pp. 2189–2202, Nov. 2012.

[16] M. Cheng, Z. Zhang, W. Lin, and P. Torr, "BING: Binarized normed gradients for objectness estimation at 300fps," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 3286–3293.

[17] 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.

[18] B. Du, Y. Zhang, L. Zhang, and D. Tao, "Beyond the sparsity-based target detector: A hybrid sparsity and statistics based detector for hyperspectral images," *IEEE Trans. Image Process.*, vol. 25, no. 11, pp. 5345–5357, Nov. 2016.

[19] J. Han *et al.*, "Efficient, simultaneous detection of multi-class geospatial targets based on visual saliency modeling and discriminative learning of sparse coding," *ISPRS J. Photogramm. Remote Sens.*, vol. 89, no. 1, pp. 37–48, 2014.

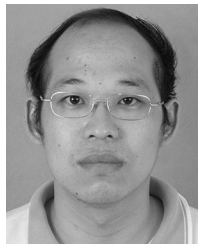
[20] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in *Proc. IEEE Int. Conf. Comput. Vis.*, Oct. 2003, pp. 1470–1477.

[21] F. Zhang, B. Du, and L. Zhang, "Saliency-guided unsupervised feature learning for scene classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2175–2184, Apr. 2015.

[22] B. Du, W. Xiong, J. Wu, L. Zhang, L. Zhang, and D. Tao, "Stacked convolutional denoising auto-encoders for feature representation," *IEEE Trans. Cybern.*, vol. PP, no. 99, pp. 1–11, 2016, doi: 10.1109/TCYB.2016.2536638.

[23] J. Han, D. Zhang, G. Cheng, L. Guo, and J. Ren, "Object detection in optical remote sensing images based on weakly supervised learning and

high-level feature learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 6, pp. 3325–3337, Jun. 2015.
[24] J. Zhang, X. Lin, Z. Liu, and J. Shen, "Semi-automatic road tracking by template matching and distance transformation in urban areas," *Int. J. Remote Sens.*, vol. 32, pp. 8331–8347, 2011.
[25] S. An, P. Peursum, W. Liu, and S. Venkatesh, "Efficient algorithms for subwindow search in object detection and localization," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 264–271.
[26] G. Cheng, L. Guo, T. Zhao, J. Han, H. Li, and J. Fang, "Automatic landslide detection from remote-sensing imagery using a scene classification method based on BoVW and pLSA," *Int. J. Remote Sens.*, vol. 34, no. 1, pp. 45–59, 2013.
[27] Z. Li and L. Itti, "Saliency and gist features for target detection in satellite images," *IEEE Trans. Image Process.*, vol. 20, no. 7, pp. 2017–2029, Jul. 2011.
[28] F. Wang, W. Zuo, L. Zhang, D. Meng, and D. Zhang, "A kernel classification framework for metric learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 9, pp. 1950–1962, Sep. 2015.
[29] 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, Sep. 2010.
[30] G. Cheng, J. Han, P. Zhou, and L. Guo, "Multi-class geospatial object detection and geographic image classification based on collection of part detectors," *ISPRS J. Photogramm. Remote Sens.*, vol. 98, no. 1, pp. 119–132, 2014.
[31] Z. Shi, X. Yu, Z. Jiang, and B. Li, "Ship detection in high-resolution optical imagery based on anomaly detector and local shape feature," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 8, pp. 4511–4523, Aug. 2014.
[32] E. Li, J. Femiani, S. Xu, X. Zhang, and P. Wonka, "Robust rooftop extraction from visible band images using higher order CRF," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 8, pp. 4483–4495, Aug. 2015.
[33] G. E. Hinton, S. Osindero, and Y. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006.
[34] G. Cheng, P. Zhou, and J. Han, "Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7405–7415, Dec. 2016.
[35] J. Yin, H. Li, and X. Jia, "Crater detection based on gist features," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 8, no. 1, pp. 23–29, Jan. 2015.
[36] A. O. Ok, C. Senaras, and B. Yuksel, "Automated detection of arbitrarily shaped buildings in complex environments from monocular VHR optical satellite imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 3, pp. 1701–1717, Mar. 2013.
[37] D. Zheng, S. Yu, and L. Yun, "The optimization model of target recognition based on wireless sensor network," *Int. J. Distrib. Sensor Netw.*, vol. 2014, pp. 1–9, 2014.
[38] Y. Zhang and D. Yeung, "Transfer metric learning by learning task relationships," in *Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2010, pp. 1199–1208.
[39] M. Zhou *et al.*, "Nonparametric Bayesian dictionary learning for analysis of noisy and incomplete images," *IEEE Trans. Image Process.*, vol. 21, no. 1, pp. 130–144, Jan. 2012.
[40] J. Hoffman, E. Rodner, J. Donahue, K. Saenko, and T. Darrell, "Efficient learning of domain-invariant image representations," in *Proc. ICLR*, 2013, pp. 1–9.
[41] J. Yang, R. Yan, and A. G. Hauptmann, "Cross-domain video concept detection using adaptive SVMS," in *Proc. ACM Int. Conf. Multimedia*, 2007, pp. 188–197.
[42] R. Achanta, F. Estrada, P. Wils, and S. Susstrunk, "Salient region detection and segmentation," in *Proc. Int. Conf. Comput. Vis. Syst.*, 2008, pp. 66–75.
[43] C. Chang and C. Lin, "Libsvm: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–27, 2011.
[44] C. Lv and Q. Zhao, "A universal PCA for image compression," in *Proc. Int. Conf. Embedded Ubiquitous Comput.*, 2005, pp. 910–919.
[45] H. S. Prasantha, H. L. Shashidhara, and K. N. B. Murthy, "Image compression using SVD," in *Proc. Int. Conf. Comput. Intell. Multimedia Appl.*, 2007, vol. 3, pp. 143–145.
[46] Z. Yuan and E. Oja, "Projective nonnegative matrix factorization for image compression and feature extraction," in *Proc. Scand. Conf. Image Anal.*, 2005, pp. 333–342.
[47] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. Berlin, Germany: Springer-Verlag, 2006.



Chunlei Huo (M'11) received the B.S. degree in applied mathematics from Hebei Normal University, Shijiazhuang, China, in 1999, the M.S. degree in applied mathematics from Xidian University, Xi'an, China, in 2002, and the Ph.D. degree in pattern recognition and intelligent system from the Institute of Automation, Chinese Academy of Sciences, Beijing, China, in 2009. He is currently an Associate Professor in the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences. His current research interests include remote sensing image processing, computer vision, pattern recognition, so on.

Zhixin Zhou is currently a Professor in Beijing Institute of Remote Sensing. His research interests include computer vision, pattern recognition, and remote sensing.

Kun Ding received the B.S. degree in automatic control from Tianjin University of Science and Technology, Tianjin, China, in 2011 and the M.S. degree in pattern recognition and intelligent system from the Institute of Automation, Chinese Academy of Sciences, Beijing in 2014, where he is currently working toward the Ph.D. degree. His current research interests include pattern recognition and machine learning.

Chunhong Pan received the B.S. degree in automatic control from Tsinghua University, Beijing, China, in 1987, the M.S. degree from the Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Shanghai, China, in 1990, and the Ph.D. degree in pattern recognition and intelligent system from the Institute of Automation, Chinese Academy of Sciences, Beijing, in 2000. He is currently a Professor in the National Laboratory of Pattern Recognition of Institute of Automation, Chinese Academy of Sciences. His research interests include computer vision, image processing, computer graphics, and remote sensing.