International Journal of Signal Processing, Image Processing and Pattern Recognition
Vol.9, No.3 (2016), pp.341-354
http://dx.doi.org/10.14257/ijsip.2016.9.3.30

A Novel Fast Remote Sensing Targets Detection Model based on KSC-SBOW

Mengxi Xu[1,2], Quansen Sun[1], Yingshu Lu[3], Fengchen Huang[3] and Chenrong Huang[2]

[1]*School of Computer Science and Technology, Nanjing University of Science & Technology, Nanjing 210094, China.*
[2]*School of Computer Engineering, Nanjing Institute of Technology, Nanjing 211167, China.*
[3]*College of Computer and Information, Hohai University, Nanjing 211100, China.*
E-mail: mxxu26@126.com

Abstract

Automatic detection for remote sensing targets is a profound challenging problem in the field of remote sensing image analysis. This paper presents a novel fast targets detection model based on kernel sparse coding with spatial bag of visual words (KSC-SBOW) in the optical remote sensing images. Before the implementation of the sliding window for targets detection, a saliency prediction model is introduced to predict the suspected targets which can extremely depress the influence of backgrounds and rationally decrease the computational cost. Following the determination of a processing sliding window by the saliency prediction model and the SIFT features extraction from the image patch, the kernel sparse coding is utilized to encode the features for a lower reconstruction error and the spatial information is added by the spatial-pyramid mapping method in the KSC-SBOW description model. Specifically, we propose the principal component analysis method based kernel orthogonal matching pursuit algorithm (KPOMP) to solve the problem of kernel sparse coding. In KPOMP, histogram intersection kernel works as the measurement kernel to more effectively capture the similarities among those Scale Invariant Feature Transform (SIFT) features and the principal component analysis method is implemented for the kernel dimensionality reduction to speed up the coding process with the guarantee of coding effectiveness. Finally, the KSC-SBOW model is combined with linear support vector machine for the targets detection. In a number of targets detection experiments, we demonstrate that the proposed model achieves outstanding performance in terms of the detection accuracy and operating rate.

Keywords*: Targets detection, Saliency prediction, Kernel sparse coding, Spatial bag of visual words, Kernel dimensionality reduction*

1. Introduction

Automatic targets detection is one of the most important applications in the field of remote sensing image (RSI) analysis, and still a challenging task as it is difficult to distinguish target of some types from others in RSI. In the computer vision, the model of "Bag of Visual Words" (BOW) has been widely researched and applied in images classification and recognition tasks (*i.e.*, locality-constrained linear coding, super vector coding, spatial-pyramid matching using sparse coding, *etc*.) [1-3].

Recently, BOW has been introduced into the domain of remote sensing targets detection and achieved brilliant performance. For instance, work by Xu et al. [4] uses BOW and support vector machine (SVM) for object classification in RSIs. In [5], Sun et al. proposes a new structure information representation combined with

ISSN: 2005-4254 IJSIP

spatial-sparse-coding-BOW (SSCBOW) to detect aircraft target by linear SVM in high resolution RSI. Wu *et. al.,* [6] using the structure of the BOW model presents a novel direction estimation method combined with their jigsaw matching pursuit to recognize aircraft target. However, these approaches can neither predict suspected targets (determinate the region of interests) before their detections which means redundant meaningless computational cost nor make great efforts to effectively capture the similarities among the extracted features in the coding phase which has negative impacts on the detection accuracy of their models. More importantly, the BOW based targets detection models mentioned above are all designed for a given type of target that limits the range of their applications.

In the meantime, works of [7-9] show that the saliency map method can be applied to predict locations of the potential candidate targets because the targets are generally distinctive from the contextual background. Moreover, Gao *et. al.*, [10, 16] proposed kernel sparse representation, which implicitly maps the input data into high-dimensional feature space through the specific kernel functions, can simultaneously ensure the data in an input space become sparser that is beneficial to improve classification accuracy [11] and remedy the drawbacks of sparse representation mentioned in [12].

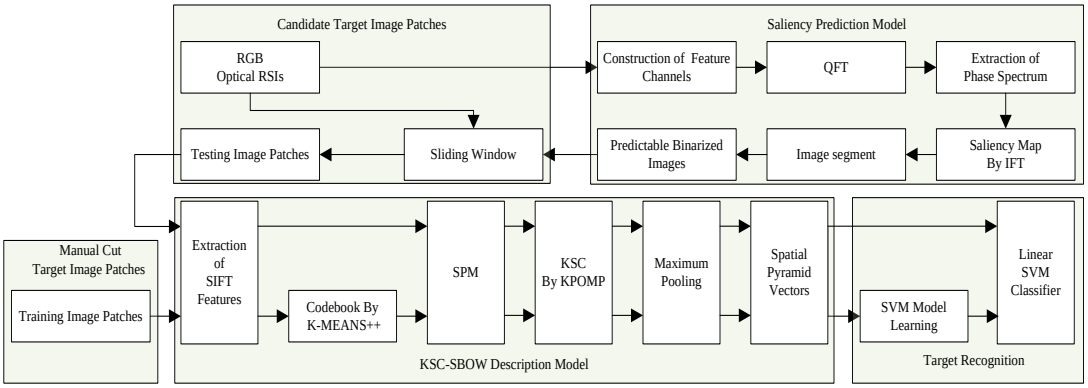


Figure 1. Framework of our Proposed Fast Targets Detection Model

Motivated by these works above, we integrate a saliency prediction model and a novel kernel sparse coding (KSC) based spatial bag of visual words (KSC-SBOW) to design our fast targets detection model shown in Figure 1 in the optical RSIs for remedying the drawbacks of the mentioned BOW based targets detection models. In the saliency prediction model, phase spectrum quaternion of Fourier transform (PQFT) algorithm [13] for the creation of saliency map is combined with image segmentation technique to segment target candidate areas from the optical RSIs. Those regions in the corresponding binarized RSI whose sum of positive pixels are greater than a threshold are retained as the potential target areas. Furthermore, in our proposed KSC-SBOW, with the aim of speeding up the feature coding process, the SIFT features [14] extracted from the training image patches are clustered for the fixed codebook by K-MEANS++ algorithm [21] rather than using the conventional KSC methods which either directly use all the primary descriptors [3] or randomly select k (manually set) features [15] to iteratively generate and optimize the codebook, because the demonstration in [16] showed that the codebook of KSC initialized with some heuristics can still achieve excellent performance. We subsequently propose the principal component analysis method based kernel orthogonal matching pursuit algorithm (KPOMP) for the problem of KSC which replaces the less effective K-MEANS vector quantization (VQ) method [17] to fast and more effectively encode the SIFT features. Specifically, we adopted histogram intersection kernel (HIK) to be measurement kernel of the feature similarities in KPOMP as Wu *et. al.,* and Maji *et.*

al., [18, 19] illustrated that HIK is more effective than Euclidean distance based kernels for the histogram features (i.e. SIFT, HOG, SURF, *etc*.). In addition, spatial pyramid mapping method (SPM) [20] is utilized in KSC-SBOW to add spatial information for the representations of image patches. After the description of image patches, linear SVM is chosen as the classifier for the target recognition tasks.

The rest of paper is organized as follows. Section 2 introduces the framework of the components in our proposed targets detection model in detail. The experimental results and evaluations are present in section 3. The conclusions are given in the last section.

2. The Proposed Targets Detection Model

The proposed targets detection model mainly contains three components: the saliency prediction model, our KSC-SBOW description model and target recognition strategy. Notably, the training target image patches are manually cut from the optical RSIs for the model learning.

2.1. Saliency Prediction Model and Candidate Targets

In the targets detection model, the target image patches which are to be detected, are obtained by the sliding window implemented on the testing RSIs. While the conventional targets detection methods based on BOW model (*i.e.*, [5, 6]) are all to directly use a sliding window to scan the whole testing large-scale RSI and take each window as the image patch to be detected. This type of models fail to provide the target candidates for the subsequent targets detection, in other words, to learn the region of interests (ROI), resulting in the redundant computational cost.

Fortunately, recently, Guo *et. al.*, [13] has proposed the phase spectrum quaternion of Fourier transform (PQFT) algorithm to create saliency map to predict human fixations for their detection model applied in image and video compression and achieved excellent performance. Inspired by this work and the outperformance of the prediction framework with the saliency method for the automatic targets detection model in the high-resolution remote sensing image by Han *et. al.*, [22], we introduce a saliency prediction model which adopts PQFT algorithm to create the saliency maps and performs image segment technique on the saliency maps to obtain the corresponding predictable binarized images, to determine the ROIs for the following process.

In the saliency prediction model, as shown in Figure 1, the PQFT algorithm mainly has three main steps as follows. First of all, it splits a RGB RSI into a four independent feature channels (two color channels, one intensity channel, and one motion channel) to capture color, intensity, and motion features, respectively. To be specific, in our model, the motion feature is set to zero as what we deal with is the static optical RSIs. After that, the quaternion Fourier transform (QFT) representation of an image with four feature channels is calculated to extract its phase spectrum. Finally, inverse Fourier transform (IFT) is utilized to reconstruct the four feature channels image as the saliency map of the RSI. As a result, its corresponding predictable binarized segmentation can be obtained by the image segment technique in which the segment threshold is computed as follows:

$$Threshold = Mean + n \times Variance, \quad (1)$$

where *Mean* and *Variance* is the mean and variance value of the saliency map intensity, respectively, and n is experimentally set to be 20.

Therefore, when a sliding window is performed on an optical RSI, if the sum of its contained positive pixels in the corresponding binarized image is larger than T (experimentally set to be 800 in our detection model), the image patch obtained by the sliding window would be regarded as a given candidate target. If not, the image patch would be abandoned.

2.2. The Proposed KSC-SBOW Model

BOW originating from the domain of text document classification, has been introduced into computer vision for about ten years. BOW can be applied to describe images (or image patches) and generate unified image representations. It mainly contains three steps [18]: SIFT features are first extracted from the image patches, then, these features are quantized using pre-generated codebook. Finally, the frequency histogram of visual words in codebook for image representation is computed. The codebook is generated by the K-MEANS algorithm which clusters the SIFT features of training images to be *K* visual words.

However, the conventional BOW model above, on one hand, unconsciously discards spatial information of local features from images which means its incapability of exact matching, on the other hand, uses the K-MEANS VQ method, which is a hard assignment method assigning the nearest visual word to a SIFT feature as the feature code, resulting in severe information loss.

2.2.1. Spatial Pyramid Mapping: In order to make full use of the geometric information of local features in the image patches, SPM is utilized in our proposed KSC-SBOW description model to add spatial information for the target representations.

In the framework of SPM, one image is divided into increasingly sub-regions in the increasing spatial-pyramid layers, and we adopt the top three layers in which each image is divided into 1×1, 2×2, 4×4 sub-regions in increasingly layers, respectively. Suppose that the size of codebook is set to be *K*, then a SIFT feature in an image patch can be encoded as $v \in R^{1\times(21K)}$ after the implementation of SPM method and coding phase.

2.2.2. Kernel Sparse Coding: With the aim to reduce information loss in the feature coding phase by the restrictive constraints of K-MEANS VQ, the ScSPM model [3] adopt a looser constraints known as the sparse coding algorithm (SC) to learn a more sparse and discriminative feature coding. SC provides a class of algorithms for finding succinct representations of stimuli, and is capable of discovering basis functions that capture higher-lever features in data [23]. Sparse coding algorithm which has soft constraints (relaxed constraints on v_n) can be seen as an extensive version of K-MEANS framework.

Suppose that the SIFT features of training images is set as $X = \left[x_1, x_2, x_3, \cdots, x_N\right] \in R^{N\times D}$ and $U = \left[u_1, u_2, u_3, \cdots, u_K\right] \in R^{K\times D}$ is the codebook. SC algorithm is to sparsely and linearly reconstruct the input data, namely, $x_m = v_1u_1 + v_2u_2 + v_3u_3 + \cdots + v_Ku_K$. In order to generate sparse coding codebook, the problem can be formulated as the following equation:

$$\min_{V,U} \frac{1}{2}\sum_{n=1}^{N}\left\|x_n - v_nU\right\|_2 + \lambda\left\|v_n\right\|_1$$

$$s.t.\ \left\|d_k\right\| \le 1, \forall k = 1,2,3,\cdots,K, \tag{2}$$

where $\lambda \ge 0$ a constant parameter for sparsity, and $\left\|\bullet\right\|_1$ is the *L*1-norm. Notably, the codebook *D* is over-complete, i.e. *K*>*D*. After that, given a codebook, the objective of it can be rewritten as the follows:

$$\min_{v} \frac{1}{2}\left\|x - vU\right\|_2 + \lambda\left\|v\right\|_1, \tag{3}$$

the construction error is the first term of equation (4), and the rest term is to control the sparsity of the coefficients v. This optimization problem is well known as LASSO in the statistical literature which is a linear regression problem with $L1$ norm regularization on the coefficients.

However, as Lee *et. al.*, [23] pointed out that learning large, highly over-complete sparse representations is extremely expensive and Wang *et. al.*, [22] showed that the codebook initialized with some heuristics can still achieve excellent performance, we propose to use the K-MEANS++ algorithm to cluster the centers as the fixed codebook U for largely reducing the computational cost. Moreover, due to the nonlinear generalization performance of kernel trick and success in computer vision applications, kernel sparse coding is proposed to get the sparse representation for a mapped feature using the mapped basis in the high-dimensional space and outperforms the SC algorithm used in BOW for image classification [16]. In that way, we propose to adopt the KSC algorithm to replace the equation (3) in the coding phase to capture more effective feature codes.

Suppose that $\varphi(\bullet)$ is set as an implicit mapping function which maps feature x into a kernel feature space. In that way, $\varphi(U) \in R^{N\times H}$ and $\varphi(x) \in R^{N\times H}$ ($H \gg D$) are used to replace the codebook U and input feature x in the high-dimensional kernel feature space F, respectively, then the equation (3) can be rewritten as follows:

$$J(v) = \min_{v} \frac{1}{2}\left\|\varphi(x) - v\varphi(U)\right\|^2 + \lambda\left\|v\right\|_1, \tag{4}$$

and it can be solved by the equation (5),

$$\hat{v} = \arg\min_{v}\left\|v\right\|_1 \text{ subject to } \left\|\varphi(x) - v\varphi(U)\right\|_2 \le \varepsilon. \tag{5}$$

It seems that the equation (5) can be solved by directly performing the orthogonal matching pursuit algorithm (OMP) [24]. While Zhang *et. al.*, [25] argued that it is not practical to directly solve the optimization problem (5), that is because, on one hand, if F is known, the computational complexity of (5) is much larger than that of equation (4) due to $H \gg D$, on the other hand, if F is unknown, we can not explicitly obtain the $\varphi(U)$ and $\varphi(X)$. Fortunately, we can access F by the kernel trick, so that the equation (5) can be modified as equation (6),

$$\hat{v} = \arg\min_{v}\left\|v\right\|_1 \text{ subject to } \left\|\varphi(x)\varphi(U)^T - v\varphi(U)\varphi(U)^T\right\|_2 \le \xi, \tag{6}$$

and (6) is equal to

$$\hat{v} = \arg\min_{v}\left\|v\right\|_1 \text{ subject to } \left\|Y - vQ\right\|_2 \le \xi, \tag{7}$$

where $K(\cdot,\cdot)$ is a Mercer kernel function and $Y = (K(x,U))_{1\times K} = \varphi(x)\varphi(U)^T$, $Q = (K(u_i,u_j))_{K\times K} = \varphi(U)\varphi(U)^T$.

2.2.3. The Proposed KPOMP Algorithm for KSC: If OMP is directly performed to solve the problem (7), we will find that the mapped codebook kernel matrix $Q \in R^{K\times K}$ is a symmetric positive definite matrix which means that the coefficient v is not sparse any more as the number of basis of codebook should be larger than the dimensionality of the basis in OMP algorithm for SC-like problems. What's worse, since the number of

SIFT features extracted from an image patches usually is hundreds or even thousands, the coding speed influenced by the dimensionality of basis and the number of feature, has great impacts on the speed of the targets detection model. Therefore, performing dimensionality reduction (DR) for Q and Y is of necessity and [25] experimentally showed that principal component analysis method (PCA) is an effective DR method in KSC-like algorithms. Therefore, we propose KPOMP algorithm, which consists of PCA for the kernel DR and OMP for the coefficient v in the kernel feature space, to fast and effectively solve the problem of (7).

In the proposed KPOMP, we set a DR matrix of PCA as $B = [\beta_1, \beta_2, \beta_3, \ldots, \beta_h] \in R^{h \times K}$ which are h normalized eigenvectors of the codebook kernel matrix Q ($h = \lfloor p \times K \rfloor$, $0 < p \leq 1$, where $\lfloor x \rfloor$ is the function to get the nearest positive integer of x). After that, the equation (7) with DR can be rewritten as follows:

$$\hat{v} = \arg\min_{v} \|v\|_1 \text{ subject to } \left\|YB^T - vQB^T\right\|_2 \leq \xi, \tag{8}$$

then performing OMP to solve the problem of (8) is our proposed KPOMP shown in algorithm 1.

To be specific, recent works [16, 18, 20] have demonstrated that HIK is a more effective measurement than Euclidean distance for histogram based features, we adopt HIK for SIFT features in our proposed KPOMP algorithm to obtain more informative and discriminative representations.

In short, the proposed KPOMP algorithm can be summarized as follows:

Step 1: Set the recovery residual threshold $0 < \chi \leq 1$, the sparsity upper bound $0 < \eta \leq K$ for the OMP algorithm and the size of the reduced dimensionality $0 < h \leq K$.

Step 2: Calculate the codebook kernel matrix $Q \in R^{K \times K}$ and its corresponding normalized eigenvectors $B \in R^{h \times K}$; For each feature x, calculate its kernel vector $Y \in R^{1 \times K}$.

Step 3: Reduce the dimensionality of Q and Y by multiplying the matrix B, set the results as $A = QB^T = \left[a_1, a_2, a_3, \cdots, a_K\right] \in R^{K \times h}$ and $y = YB^T \in R^{1 \times h}$, respectively.

Step 4: Implement the OMP algorithm with the new codebook A and input vector y to get the sparse coefficient $v \in R^{1 \times K}$ as the image representation.

2.2.4. Maximum Pooling: Following [5, 16], maximum pooling method which is well established by biophysical evidence in visual cortex and empirically illustrated by many image classification algorithms [3], is used to pool the feature codes of an image in our KSC-SBOW model. Suppose that an image (or an image patch) has *M* SIFT features, after the implementation of SPM and KSC, the image can be represented by the coefficient matrix $V = \left[v_1, v_2, v_3, \cdots, v_M\right] \in R^{M \times (21K)}$. Therefore, in the pooling procedure, the maximum pooling function can be formulated as:

$$Max: r = \max\left(|v_1|, |v_2|, |v_3|, \cdots, |v_M|\right), \tag{9}$$

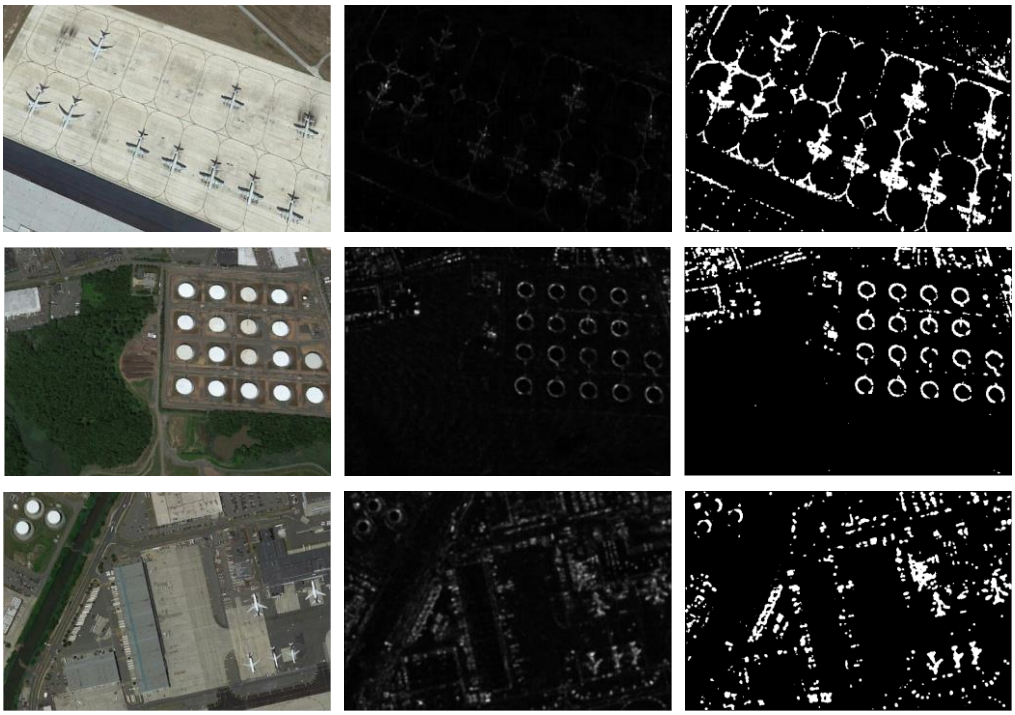
where $r \in R^{1 \times (21K)}$ is the unified spatial-pyramid coefficient vector for an image representation.

2.3. Target Recognition

After the description of the candidate target image patches, linear SVM with linear kernel is chosen as our classifier for the target recognition tasks and the non-maximum suppression method in [5] is adopted for the problem of multiple overlapped detection windows of the same target in our targets detection model.

3. Experiments and Evaluations

We evaluate the proposed fast targets detection model on 150 optical RSIs collected from the publicly accessible Google Earth, as shown in Figure 2 (a). 100 image patches of aircrafts, 100 image patches of storage tanks, and 300 image patches of background shown in Figure 3, are manually cut from the optical RSIs for the training phase. In the detection procedure, before performing the sliding window on the RSIs, the saliency prediction model is implemented to segment created the corresponding saliency maps to get the predictable binarized images, shown in Figure 2 (b) and (c), respectively. Notably, the valid detection is defined as a sliding window contains more than 60% part of the given target image.



(a) (b) (c)

Figure 2. A Number of Detection Samples. (a) Experimental Optical RSIs; (b) The Corresponding Saliency Maps; (c) The Segmented Predictable Binarized Images

In our proposed KPOMP, we set the recovery residual threshold ξ=0.001, the sparsity upper bound $\eta = \lceil 0.05K \rceil$ and the proportion of dimensionality reduction p=0.5. For the recognition, linear SVM using linear kernel is adopted as the classifier in the model. In order to evaluate the performance of our proposed methods, we take the recall-precision curve (RPC) to be the measurement, where *Precision* and *Recall* are defined as follows:

$Precision=TP / (TP+FP)$

$$Recall=TP / (TP+FN), \quad (10)$$

the equation (10) refers to that *Precision* is to compute the value which is the number of true positives divided by the number of all detected elements and *Recall* is the number of true positives divided by the number of actual target. The best performance in PRC curve is that the method yields both the best *Precision* and *Recall*. While the detection model may obtain a high *Recall* value but a bad *Precision* or a high *Precision* value but a bad *Recall*. To evaluate the *Precision* and *Recall* results, we adopt *F1-measure* (*F1*) [22] as the metric in which the optimal performance in terms of *Precision* and *Recall* is defined as the one leading to the highest *F1*. In addition, *F1-measure* can be computed as: $F1\text{-}measure = (2 \times Precision \times Recall) / (Precision + Recall)$.

(a) (b) (c)

Figure 3. Samples of Training Target Image Patches. (a) Aircraft Image Patches; (b) Storage Tank Image Patches; (c) Background Image Patches

3.1. Targets Detection Results

Three BOW based image description models, BOWSVM, KSPM [20] and our proposed KSC-SBOW, combined with our predictable detection framework respectively are implemented for the targets detection using the same training data and testing data. BOWSVM is the conventional BOW using linear SVM with linear kernel and KSPM is the nonlinear kernel SPM using spatial-pyramid histograms and the pyramid match kernel (PMK).

The codebook sizes of these methods are all set to be 500. For the sliding window, the side length of it is from 40×40 to 160×160 pixels with the interval of 10 pixels and the sliding step is set to be 25% of its side length. The samples of experimental detection results are shown in Figure 4.

From Figure 5, we can get that our proposed KSC-SBOW method outperforms the other methods both in terms of *Precision* and *Recall*. With the same *Recall*, the *Precision* of KSC-SBOW is better than the ones of KSPM and BOWSVM methods which means our KSC-SBOW method can achieve the lowest alarm false rate with the same number of true positives. And with the same *Precision*, the *Recall* of KSC-SBOW method is the best one in the three methods which means the proposed method can detect the largest number of true positives with the same false alarm rate. Moreover, due to the implementation of the saliency prediction method, the *Precision* results always keep large value when given a low *Recall* result which is to say that our saliency method can ensure a high valid prediction rate resulting in a low false alarm rate. Therefore, we can conclude that the proposed KSC-SBOW method remarkably improve the performance of the targets detection model. In addition, from the above experiments, we find that the side length of the sliding window which is set to be 80 pixels obtains both high *Precision* and *Recall* in the KSC-SBOW, and the following experiments all adopt 80 pixels for the side length of

the sliding window.

(a) (b) (c)

Figure 4. Targets Detection Results. (a) Results of BOWSVM based Detection Model; (b) Results of KSPM based Detection Model; (c) Results of our Detection Model

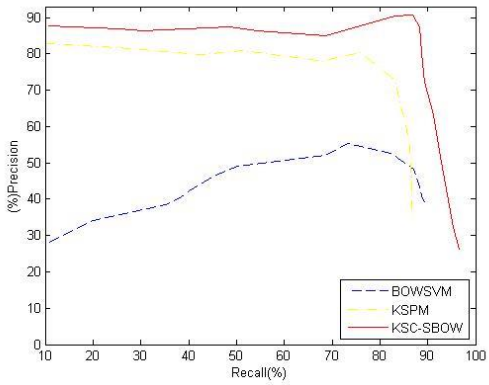


Figure 5. PRCs of the KSC-SBOW, KSPM and BOWSVM methods.

3.2. Our Proposed Model Revisit

The proposed targets detection model was implemented on a Windows PC platform using MATLAB 2013a software with 2GB RAM and Intel Core i3 2.27GHz. The measurement *Mean Time* is added to evaluate the average running time of our algorithm, and its definition will be present in the experiments, specifically.

3.2.1. Kernel Dimensionality Reduction in KPOMP: We implement different dimensionality reduction of the codebook kernel matrix by the PCA method in our proposed KPOMP algorithm, the proportion p is set to be 0.1, 0.3, 0.5, 0.8 and 1 (p=1 means no dimensionality reduction), respectively. To evaluate the speed of our proposed KPOMP algorithm, we test all the training samples and compute *Mean Time* of the image description

process which is the average running time cost by translating an image into a spatial-pyramid vector using KPOMP in our KSC-SBOW description model.

In Figure 6, it is appearant that when the proportion p is set to be 0.5, the detection model achieves the best *Precision* and relatively better performance of the corresponding *Recall*, notably, the speed of the case, measured by *Mean Time* shown in Table 1, is faster than the case using the codebook kernel matrix without the dimensionality reduction. Whatsmore, though the dimensionality of the codebook kelnel matrix has been reduced to 10%, the performance is much outstanding. In contrast with *Precision* steep curve, the *Recall* has a smooth curve which means that the KPOMP algorithm can keep the number of true positives and simultaneously reduce the number of missrecognitions. Therefore, we can conclude that our proposed KPOMP algorithm have the ability to capture the main discriminative information and depress the disturbing noises for the kernel matrix of codebook and thanks to the large dimensionality reduction, the KPOMP algorithm becomes faster in speed and more effective. Due to the highest *F1-measure* and relatively lower *Mean Time*, the concludision can be obtained that when the proportation of kernel dimensionality reduction p is set to be 0.5, our targets detection model outperforms the others in terms of both recogition rate and running time. Specifically, the following sections all adopt p=0.5 for the experiments except special declarations.

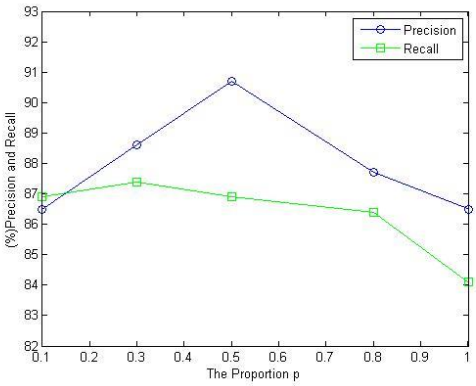


Figure 6. *Precision* and *Recall* Results of our Fast Target Detecion Model using Different Proportion of Dimensionality Reduction in the KPOMP Algorithm

Table 1. *F1-measure* and *Mean Time* Results of Describing a Target Image using Different Proportion of Dimensionality Reduction in the KPOMP Algorithm

p	0.1	0.3	0.5	0.8	1
F1	0.867	0.875	**0.888**	0.870	0.855
Mean Time(s)	**0.112**	0.361	0.545	0.843	1.020

3.2.2. Influence of Kernels: In our targets detection model, two Mercer kernels are used in the proposed KPOMP algorithm and SVM classifier. Linear kernel (LK), Gaussian kernel (GK) and HIK are respectively utilized in the KPOMP algorithm, and HIK, PMK and LK are separately used in the SVM classifier. The selection of the kernels has great impacts on the performance of our targets detection model. While the selection may obtain a high *Recall* value but a bad *Precision* which means the selected kernels lead to a high false alarm rate with a good *Recall*. Therefore, the experiments are implemented to evaluate the performance of selected couples of kernels in terms of *Precision*, *Recall* and *F1-measure* and the results are shown in Table 2. The targets detection model using HIK in the KPOMP algorithm and LK in SVM classifier, namely HIK-LK, achieves the

highest *F1* in which the *Precision* and *Recall* both have good results, which means HIK-LK has the ability to ensure a lowest false alarm rate with a large number of true positives.

As the result, we can conclude that the HIK is more effective than the Euclidean distance based kernels for SIFT features in the feature coding process and it combined with SVM using lineal kernel can obtain the best performance of our proposed targets detection model.

Table 2. Performance Comparisons of Different Selections of Two Kernel in Terms of *Trecision*, *Recall* and *F1-measure*.

KOMP	LK			GK			HIK		
SVM	*Precision*	*Recall*	*F1*	*Precision*	*Recall*	*F1*	*Precision*	*Recall*	*F1*
HIK	76.4%	92.1%	0.835	87.6%	79.4%	0.833	74.1%	**93.4%**	0.826
PMK	75.9%	86.9%	0.810	81.5%	88.7%	0.849	73.3%	89.7%	0.807
LK	86.9%	84.1%	0.855	79.7%	91.6%	0.852	**90.7%**	86.9%	**0.888**

3.2.3. Comparison of Pooling Methods: Following [3, 26, 27], we evaluate three pooling methods, namely, the mean of absolute values (*Abs*), the square root of mean squared statistics (*Sqrt*) and maximum pooling (*Max*), as follows:

$$Abs: r = \frac{1}{M}\sum_{i=1}^{M}\|v_i\|,$$

$$Sqrt: r = \sqrt{\frac{1}{M}\sum_{i=1}^{M}v_i^2}, \quad (11)$$

To be specific, the proportion of reduced dimensionality by KPCA is set to be 1 that is equal to using the original codebook kernel matrix for the KPOMP algorithm. The results in Table 3 shows that the maximum pooling method achieves the largest *Precision* and *Recall* in three method which means *Max* is the most suitable method for the features pooling procedure in our KSC-SBOW. The reason is that maximum pooling method is well established by biophysical evidence in visual cortex and empirically illuminated by many image classification algorithms [3], and in the KSC-SBOW, due to the sparse coefficient matrix of an image generated by KPOMP algorithm, the mean algorithm based pooling methods discarding the discriminative information are unable to obtain the effective expressive image representations.

Table 3. *Precision* and *Recall* Results of our Fast Targets Detection Model using *Abs*, *Sqrt*, *Max* Pooling Method in the KSC-SBOW, Respectively

Pooling Methods	*Precision*	*Recall*
Abs	73.5%	79.2%
Sqrt	78.4%	81.3%
Max	**86.5%**	**84.1%**

3.2.4. Influence of Saliency Map: In the detection procedure, the influence of the saliency method is evaluated when the detection model adopts it. With the consideration of the computational speed and the detection precision, we adopt p=0.5 and the maximum pooling method in the contrast experiments. The *Mean Time* in Table 4 refers to a measurement which is to compute the mean time cost by the detection algorithm

performed on an optical RSI, and we test all 150 RSIs to get the average value.

Table 4 shows that the *Precision* and *Mean Time* results of the detection model without the saliency map method are much worse than the case with the saliency map method, but its *Recall* value is higher. That is because the detection model using the saliency map can predict the regions of interests and abandon those sliding windows which are thought to be the ones containing no given target, resulting to the higher *Precision* value and faster detection speed. Furthermore, the *F1-measure* results shows the better performance of the targets detection model with the saliency method. While the saliency map method has no ability to achieve perfect predictions as the existing complex background in the optical RSIs, the *Recall* value of the detection using saliency map method is rationally lower than the case using no saliency map.

Table 4. Performance of the Targets Detection using the Saliency Method and no Saliency Method in Terms of *Precision*, *Recall*, *F1-measure* and *Mean Time*

	Precision	*Recall*	*F1*	*Mean Time(s)*
Saliency Map	**90.7%**	86.9%	**0.888**	**37.632**
No Saliency Map	76.2%	**93.0%**	0.828	83.011

4. Conclusion

In this paper, we integrate the saliency prediction model and our proposed KSC-SBOW description method to design our fast targets detection model in the optical RSIs. The saliency prediction model using PQFT method and the image segment technique, is introduced to predict the suspected targets which can simultaneously accelerate the detection process and improve the detection rate. Furthermore, the KSC-SBOW description model is proposed to represent target images, in which SIFT features are extracted from the images, K-MEANS++ algorithm is adopted to generate the fixed codebook with the aim of reducing the computational cost, and the KPOMP algorithm is raised to fast and effectively solve the LASSO problem of the KSC. Moreover, the SVM classifier with linear kernel is chosen as the multi-class classifier. Finally, experimental results of the detection for the target of aircraft and storage tank in the optical RSIs demonstrate the outstanding performance of our proposed detection model in terms of detection rate and running speed.

Acknowledgement

This work is supported partly by the National Natural Science Foundation of China (No. 61273251, 61401195, 61271386), and the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (No. 13KJB520009).

Reference

[1] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang and Y. Gong, Locality-constrained linear coding for image classification, Computer Vision and Pattern Recognition, 3360-3367 **(2010)**
[2] X. Zhou, K. Yu, T. Zhang and T. S. Huang, Image classification using super-vector coding of local image descriptors, Computer Vision-ECCV, 141-154 **(2010)**
[3] J. C. Yang, K. Yu, Y. H. Gong and T. Huang, Linear spatial pyramid matching using sparse coding for image classification, Computer Vision and Pattern Recognition, 1794-1801 **(2009)**
[4] S. Xu, T. Fang, D. R. Li and S. Wang, Object classification of aerial images with bag-of-visual Words, Geoscience and Remote Sensing Letters, 7(2): 366-370 **(2010)**
[5] H. Sun, X. Sun, H. Wang, Y. Li and X. Li, Automatic targets detection in high-resolution remote sensing images using spatial sparse coding bag-of-words model, Geoscience and Remote Sensing Letters, 9(1): 109-113 **(2012)**
[6] Q. C. Wu, H. Sun, X. Sun, D. Zhang, K. Fu and H. Wang, Aircraft recognition in high-resolution optical

satellite remote sensing images, Geoscience and Remote Sensing Letters, 12(1): 112-116 **(2015)**
[7] J. Han, K. N. Ngan, M. Li and H. J. Zhang, Unsupervised extraction of visual attention objects in color images, Circuits and System for Video Technology, 16(1): 141-145 **(2006)**
[8] L. Zhang, M. H. Tong, T. K. Marks, H. Shan and G. W. Cottrell, SUN: a Bayesian framework for saliency using natural statistics, Journal of Vision, 8(7): 1-20 **(2008)**
[9] T. Judd, K. Ehinger, F. Durand and A. Torralba, Learning to predict where humans look, Computer Vision, 2106-2113 **(2009)**
[10] S. Gao, I. W. H. Tsang and L. T. Chia, Kernel sparse representation for image classification and face recognition, Computer Vision–ECCV, Springer Berlin Heidelberg, 1-14 **(2010)**
[11] M. Jian and C. K. Jung, Class-discriminative kernel sparse representation based classification using multi-objective optimization, Signal Processing, 61(18): 4416-4427 **(2013)**
[12] J. Yin, Z. Liu, Z. Jin and W. Yang, Kernel sparse representation based classification, Neurocomputing, 77(1): 120-128 **(2011)**
[13] S. Maji, A. C. Berg and J. Malik, Classification using intersection kernel support vector machines is efficient, CVPR, 1-8 **(2008)**
[14] D. G. Lowe, Distinctive image features from scale-invariant keypoints, IJCV, 60(2): 91-110 **(2004)**
[15] L. Zhang, W. Zhou and F. Z. Li, Kernel sparse representation-based classifier ensemble for face recognition, Multimedia Tools and Applications, 74(1): 123-137 **(2013)**
[16] S. H. Gao, I. W. Tsang and L. T. Chia, Sparse representation with kernels, Image Processing, 22(2): 423-434 **(2013)**
[17] O. Boiman, E. Shechtman and M. Irani, In defense of nearest-neighbor based image classification, Computer Vision and Pattern Recognition, 1-8 **(2008)**
[18] J. X. Wu and J. M. Rehg, Beyond the Euclidean distance: Creating effective visual codebooks using the histogram intersection kernel, Computer Vision, 630-637 **(2009)**
[19] S. Maji, A. C. Berg and J. Malik, Classification using intersection kernel support vector machines is efficient, CVPR, 1-8 **(2008)**
[20] S. Lazebnik, C. Schmid and J. Ponce, Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories, Computer Vision and Pattern Recognition, 2: 2169-2178 **(2006)**
[21] D. Arthur and S. Vassilvitskii, k-means++: The advantages of careful seeding, Society for Industrial and Applied Mathematics, 1027-1035 **(2007)**
[22] J. Han, P. Zhou, D. Zhang, G. Cheng, L. Guo, Z. Liu, S. Bu and J. Wu, Efficient, simultaneous detection of multi-class geospatial targets based on visual saliency modeling and discriminative learning of sparse coding, ISPRS Journal of Photogrammetry and Remote Sensing, 89: 37-48 **(2014)**
[23] H. Lee, A. Battle, R. Raina and A. Y. Ng, Efficient sparse coding algorithms, Advances in neural information processing systems, 801-808 **(2006)**
[24] Y. C. Pati, R. Rezaiifar and P. S. Krishnaprasad, Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition, Signals, Systems and Computers, 40-44 **(1993)**
[25] L. Zhang, W. D. Zhou, P. C. Chang, J. Liu, Z. Yan, T. Wang and F. Z. Li, Kernel sparse representation-based classifier, Signal Process, 60(4): 1684-1695 **(2012)**
[26] A. Shi, L. Xu, F. Xu and C. Huang, Multispectral and panchromatic image fusion based on improved bilateral filter, Journal of Applied Remote Sensing, 5(1): 053542-053542 **(2011)**
[27] F. Tang, H. Lu, T. F. Sun and X. Jiang, Efficient image classification using sparse coding and random forest, Image and Signal Processing, 781-785 **(2012)**

Authors

Mengxi Xu, She received the MSc degree from UNESCO-IHE, the Netherlands. She is currently a lecturer in School of Computer Engineering, Nanjing Institute of Technology, China. Her current research interests include image processing, information processing system and its application.

Quansen Sun, He is a professor in School of Computer Science and Engineering, Nanjing University of Science and Technology, China. He received his PhD degree from Nanjing University of Science and Technology, China. Currently, his research area includes theory and application of pattern recognition, image processing, analysis and recognition, remote sensing information intelligent

processing.

Yingshu Lu, He received B.S degree in electronic information science and technology from Jiangsu University of Science and Technology, China. He is currently a Master degree candidate in College of Computer and Information, Hohai University, China. His research interests are image processing, pattern recognition and signal processing.

Fengchen Huang, He is an associate professor in College of Computer and Information, Hohai University, China. He received the M.S. degree in Communications and Electronic System from University of Electronic Science and Technology of China. His research areas are signal processing in remote sensing and remote control, and communication systems.

Chenrong Huang, She is a professor in School of Computer Engineering, Nanjing Institute of Technology, China. She received her PhD degree from Nanjing University of Science and Technology, China. Currently, her research area includes image processing, bionic system modeling and information processing.

