

Airplane detection based on rotation invariant and sparse coding in remote sensing images

Liu Liu, Zhenwei Shi*

Image Processing Center, School of Astronautics, Beihang University, Beijing, China

Abstract

Airplane detection has been taking a great interest to researchers in the remote sensing filed. In this paper, we propose a new approach on feature extraction for airplane detection based on sparse coding in high resolution optical remote sensing images. However, direction of airplane in images brings difficulty on feature extraction. We focus on the airplane feature possessing rotation invariant that combined with sparse coding and radial gradient transform (RGT). Sparse coding has achieved excellent performance on classification problem through a linear combination of bases. Unlike the traditional bases learning that uses patch descriptor, this paper develops the idea by using RGT descriptors that compute the gradient histogram on annulus round the center of sample after radial gradient transform. This set of RGT descriptors on annuli is invariant to rotation. Thus the learned bases lead to the obtained sparse representation invariant to rotation. We also analyze the pooling problem within three different methods and normalization. The proposed pooling with constraint condition generates the final sparse representation which is robust to rotation and detection. The experimental results show that the proposed method has the better performance over other methods and provides a promising way to airplane detection.

Keywords: Airplane detection, sparse coding, rotation invariant, radial gradient transform, constraint pooling.

*Corresponding author. Tel.: +86-10-823-39-520; Fax: +86-10-823-38-798.

Email addresses: liuliubh@gmail.com (Liu Liu), shizhenwei@buaa.edu.cn
(Zhenwei Shi)

1. Introduction

Target detection in high resolution optical remote sensing images is a challenging task owing to its changing appearance and arbitrary direction. More recently, airplane detection, as an important detected target, has gained hot research and exploration [1] [2] [3] in military and civil applications, such as airfield surveillance. With the resolution growing, more spatial information are provided so that we could know more about the feature information.

The problem of airplane detection is generally considered as exploiting target feature exclusively to make decision regarding the type of each sample—target or non-target, known as binary classification. Arbitrary direction of airplane in images brings difficulty on detection. The first need is to explore a robust feature that allows the airplane to be well discriminated without the influence by rotation. We focus on the issue of features for airplane detection on sparse coding. Sparse coding, as an emerging signal processing technique, has attracted more and more researchers' attention due to its comprehensive theoretical studies [4] and excellent performance on machine learning and computer vision problems [5] [6]. The general sparse coding process consists of two-phase: dictionary learning and sparse representation. Local descriptors, such as scale invariant feature transform (SIFT) [7] descriptors or raw patches sampled from the image on a regular grid, are used to train dictionary for better fitting the data. The sparse representation uses the learned dictionary to find the best linear combination to represent the feature of the target. However, the general descriptor, such SIFT descriptor and HOG descriptor, dose not possess the rotation-invariant [8]. To obtain the rotation-invariant sparse representation, we apply radial gradient transform [8] descriptor to dictionary learning, thus the obtained sparse feature possesses the rotation-invariant property.

Several works have been done for airplane detection in the fields of remote sensing images, such as, shape-based method of circle frequency filter [9] uses the Fourier transform, and multiple segmentation [10] combining with contour information extracts candidate region. Xu *et al.* [11] apply an artificial bee colony with an edge potential to recognition. Coarse-to-fine process prior [12] is proposed by using high-level information of the shape. All these methods are based on the gray image information and ignore gradient information that is robust to the local geometric changes. Thus we consider the gradient histogram on the samples, and also use the gradient information for dictionary learning and sparse representation.

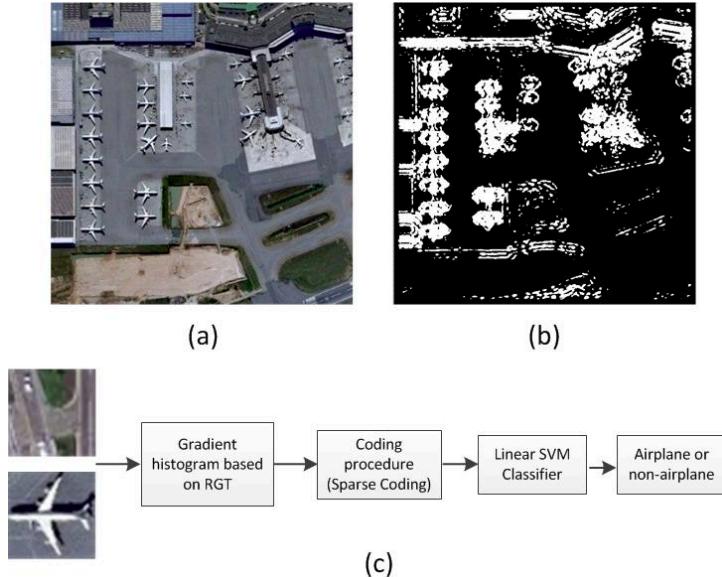


Figure 1: (a) The remote sensing image with one-meter resolution; (b) The candidate region of airplane; (c) The workflow of airplane detection based on sparse coding and radial gradient transform (RGT) descriptor through linear classification. The two small images on the left are sampled from the candidate region of remote sensing image by using sliding window

Orientation problem is the key problem in the airplane detection, because the orientation of the airplane is unpredictable within many remote sensing images. Thus we address the problem of rotation-invariant feature. Several methods have been applied to the rotation problem. Principal component analysis (PCA) method [13] estimates the main axis and uses template matching to detection; symmetry-based method [14] is to find the axis direction by minimum within-group variance dynamic threshold; and circle frequency filter [9] uses fourier transform to delete the influence of rotation. However, these methods are most based on pixel value, which could be affected by the various backgrounds of optical remote sensing images, such as illumination, shadowing, *etc.* Thus we consider the feature descriptor by using gradient information that is invariant to rotation after radial gradient transform [8].

This paper introduces a new rotation-invariant feature representation, based on sparse coding and radial gradient transform, which deals with ar-

bitrary orientation of airplane in the high-resolution optical remote sensing images. We focus on the civil airports in remote sensing images from Google map and deal with the detection of civil airplane. The civil airplane in remote sensing images, which has one-meter resolution, possesses about 40 pixels length and 40 pixels width in such images, as shown in Figure 1(a). Figure 1 (b) shows the candidate region of airplane by using circle frequency filter [9] method. The circle frequency filter could delete the rotation effect but poorly detect under complex background such that chosen as preprocess before detection. The workflow of the airplane detection is shown in Figure 1(b). The radial gradient transform is the key process on computing the sparse feature. Local descriptors are formed on annuli based radial gradient transform system that possess rotation invariant. For the sparse coding, we first train the dictionary by using local descriptors that belong to all the samples. This obtained dictionary is more effective than the unsupervised one in terms of classification. We compare three pooling methods to obtain the final sparse representation by max, average and constraint. In the airplane detection, we take linear SVM as detection model due to its linear computation complexity.

This paper is organized as follows: In Section 2, we introduce the radial gradient transform. Sparse coding methods include dictionary learning and sparse representation are presented in Section 3. Section 4 argues about the pooling methods. Detection process and experiment results are shown in Section 5 and Section 6, and concluding remarks are made in Section 7.

2. Rotation-invariant Descriptors

The orientation of airplane is various according to the situation of the airport or some other condition. It is unrealistic to train all directions of airplanes to detect airplane in remote sensing images. The reasonable method is extracting feature of airplane possessing rotation-invariant. Typical feature descriptors, such as SIFT [7] and speeded up robust feature (SURF) [15], assign an orientation to interest points before extracting descriptor. But there are not always interest points in the airplane sample. So we need an orientation invariant descriptor which eliminates the computation of finding an orientation and interpolation the relevant pixels. In this section, we mainly discuss an orientation invariant descriptor based on radial gradient transform (RGT) [8], which will be used in sparse coding section.

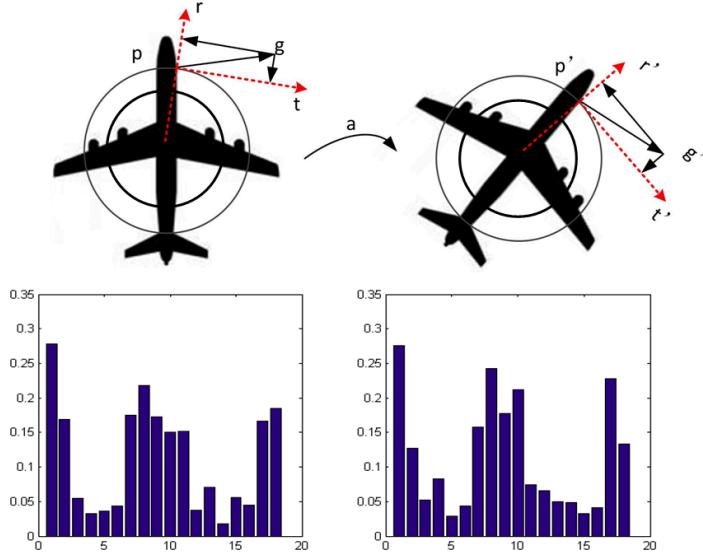


Figure 2: Illustration of radial gradients. The first line: Left: gradient \mathbf{g} is projected onto radial coordinate system (\mathbf{r}, \mathbf{t}) ; Right: the image rotates a certain angle α , the new gradient \mathbf{g}' , at the same position of airplane, projects onto new radial coordinate system $(\mathbf{r}', \mathbf{t}')$. The second line describe the gradient histogram based on annulus between two circles above. The x-coordinate is the 18 signed orientation bins; the y-coordinate is the gradient statistic information

2.1. Radial coordinate system

The general feature descriptor is based on gradient information. To make the gradient descriptor invariant to the varying orientation, we need to apply transformation to gradient information. RGT [8] projects gradient into the radial coordinate system without loss of information.

As shown in Figure 2, radial coordinate system (\mathbf{r}, \mathbf{t}) is related to the point p and the center of the image, where vector \mathbf{r} is the unit vector and its direction is from the center of image toward the point p . At the same time, unit vector \mathbf{t} is orthogonal to vector \mathbf{c} . We decompose the gradient \mathbf{g} onto radial coordinate system (\mathbf{r}, \mathbf{t}) , which obtains a new vector $(\mathbf{g}^T \mathbf{r}, \mathbf{g}^T \mathbf{t})$. Assume the airplane is rotated with a certain angle. Point p turns to point p' . The gradients of these two points are different, but the amplitudes are the same. And then project the new point p' on the new radial coordinate system $(\mathbf{r}', \mathbf{t}')$, which obtains another a new vector $(\mathbf{g}'^T \mathbf{r}', \mathbf{g}'^T \mathbf{t}')$. It is easy to verify that these two new vectors are equal:

$$(\mathbf{g}^T \mathbf{r}, \mathbf{g}^T \mathbf{t}) = (\mathbf{g}'^T \mathbf{r}', \mathbf{g}'^T \mathbf{t}') .$$

The gradient of each point on the airplane that projected on the radial coordinate system is invariant when the airplane rotates a certain angle around the center of the airplane.

2.2. Radial gradient transform descriptor

In order to obtain rotate-invariant descriptor, unlike Histograms of Oriented Gradients (HOG) [16] or SIFT [7] descriptor that computes histogram of gradient in the block, we consider the histogram of gradient in the annuli. Each point of gradient information based on radial coordinate system is invariant to rotation around the center of the example. Thus, the obtained histogram of gradient is rotation-invariant, as shown in Figure 2. The descriptors are densely sampled from the image similar to HOG descriptors. But the RGT descriptors are based on annulus around the center of the example to count the gradient information. We divide the example into different annuli, these annuli have different radius but the example statistical gradient standards, such as the number of the bins, the signed gradient direction. Our dictionary learning in the later section is based on the rotation-invariant descriptors, which is a key process on sparse coding.

3. Sparse coding

Sparse coding has been successfully applied to many fields and gained popularity among researchers working on image classification, due to its state-of-the-art performance on several benchmarks [6]. This coding refers to a general class of techniques that automatically selects a sparse set of vectors from a large pool of possible bases to encode an input feature vector. On behalf of the high-quality code book, we also use the descriptors mentioned above to train our dictionary. However, sparse feature based on the sparse coding is not rotation-invariant, because the feature descriptor is based on block and no transformational gradient. Though Yang *et al.* [17] provide translation-invariant sparse coding, it could not deal with the rotation problem. Thus, we propose such sparse coding that is invariant to rotation based on RGT descriptor and the obtained sparse feature of the airplane is robust on eliminating influence of rotation as well.

3.1. Sparse representation

Let \mathbf{X} be a set of RGT descriptors in form of annulus within an example in form of matrix, i.e. $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in R^{d \times n}$, d is the length of descriptor. Let $\mathbf{B} \in R^{d \times p}$ be a Codebook of codeword, p is the size of the codebook. The patch sparse representation is $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n] \in R^{p \times n}$. Sparse coding seeks a linear reconstruction of the given descriptor by using the bases in the dictionary. The reconstruction coefficients \mathbf{w} are sparse and are minimized by using l_1 norm to approximate the sparsest nearsolution [4]. To cater to the reconstruction error of the descriptor, the objective of sparse coding can be formulated as follow:

$$\arg \min_{\mathbf{W}} \frac{1}{n} \sum_{i=1}^n \left\{ \frac{1}{2} \|\mathbf{x}_i - \mathbf{B}\mathbf{w}_i\|^2 + \lambda \|\mathbf{w}_i\|_1 \right\}, \quad (1)$$

where λ is s lagrange multiplier. The first term in (1) is the reconstruction error, and the second term is used to control the sparsity of the sparse \mathbf{w} . Notice, the non-negative is dropped out, because of the negative \mathbf{w}_i can be absorbed by flipping the corresponding basis. Normally, the codebook B is over-complete, i. e. $p > n$. Thus the sparsity can be well reflected in capturing the salient pattern of local descriptors. For each coefficient \mathbf{w}_i , the optimization model is a linear regression problem with l_1 norm regularization and can be solved very efficiently by algorithms such as feature-sign [18].

3.2. Codebook learning

Effective image coding requires high-quality codebook \mathbf{B} . When the codebook is given, sparse representation of descriptor can be obtained. Codebook learning aims to solving the following optimization problem:

$$\begin{aligned} \min_{\mathbf{B}} L(\mathbf{W}) &= \frac{1}{n} \sum_{i=1}^n \left\{ \frac{1}{2} \|\mathbf{x}_i - \mathbf{B}\mathbf{w}_i\|^2 + \|\mathbf{w}_i\|_1 \right\} \\ \text{s.t. } &\|\mathbf{B}_i\|_1 \leq 1, \quad i = 1, 2, \dots, p, \end{aligned} \quad (2)$$

where $L(\mathbf{W})$ is loss function. For solving the codebook B, an efficient method is introduced by [18] using dual formulation. This method has the advantage of decreasing the optimization variables. Yu *et al.*[19] develope a projected Newton method to solve the optimization problem.

It is easy to see that the above objective function is the same as the one in sparse coding when given the codebook to solve the sparse representation. Sparse coding (SC) has two phases, training and coding. In training

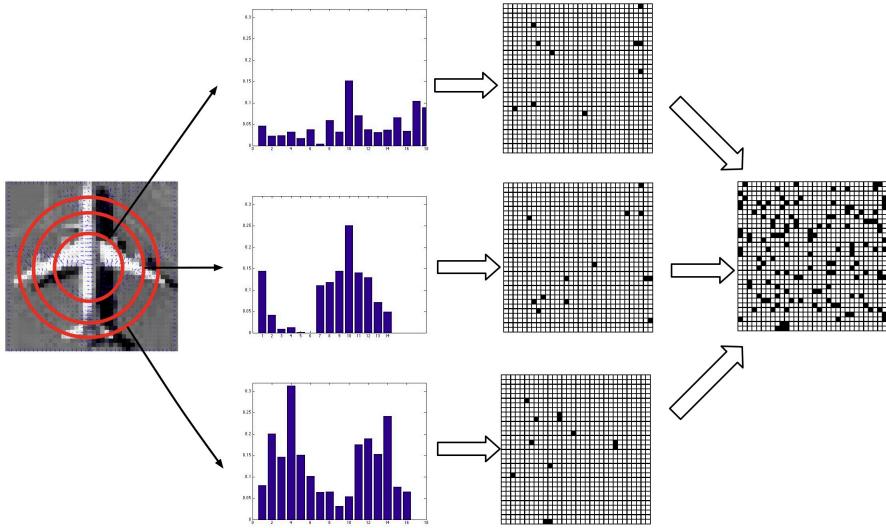


Figure 3: Illustration of sparse representation, first to obtain the gradient histogram round the annuli (left), then through sparse coding to get the sparse representation (middle), at last via pooling over these sparse vectors to obtain the final sparse representation (right).

phase, given a set of descriptor of \mathbf{X} , we can obtain codebook and sparse representation respectively by iteratively alternating optimization problem eq.(2) and eq.(1): 1) given the codebook B , compute the optimal sparse representation using efficient coding; 2) given the new coding, re-optimize the codebook. Note, we use more than 10,000 RGT descriptors from random annulus patches to train the codebook by iterating the eq.(2) and eq.(1).

4. Pooling

Pooling, which has long been an important part of recognition architecture such as convolutional network [20], gives robustness to small transformation of image. The codes of the descriptors within subregions are pooled together to form the corresponding feature [6] as the representation of image. Jia *et al.* [21] focus on the definition of receptive fields for pooling and obtain the pooled image feature by using receptive field to aggregate the activations over certain regions as global representation of the image. Boureau *et al.* [22] consider the locality in feature space to apply in object recognition. One purpose of pooling step is to produce representation that aggregates the local sparse representations without losing too much information in feature

extraction. In our paper, the pooled feature is formed by constraint optimistic. Figure 3 shows the final sparse representation process based on the middle sparse vectors. Each subregion is the annuli from the sample. The RGT descriptor corresponding to a sparse vector is obtained on the subregion. The final sparse is formed by pooling these sparses vectors to reach the final sparse representation. Here, we introduce two common poolings of mean and maximum, and the constraint pooling based on optimistic model.

4.1. Mean of absolute values(Abs)

The mean of absolute values [6] takes the average absolute values in each row of sparse vectors:

$$\mathbf{z} = \frac{1}{n} \sum_{j=1}^n |\mathbf{w}_j|, \quad (3)$$

where n is the number of sparse vectors, w_j is the j -th sparse representation of image. Before the pooled feature fed into the final classifier, it is often normalized by l_1 -norm or l_2 -norm.

4.2. Max pooling

In the method of pooling, Max pooling is to select the maximum value in each row of sparse vectors to form the feature vector to apply into categorization and detection [6][23]. The pooling function on the absolute sparse codes is following:

$$z_i = \max \{ |w_{1i}|, |w_{2i}|, \dots, |w_{ni}| \}, \quad (4)$$

where z_i is the i -th element of \mathbf{z} , w_{ij} is the i -th row and j -th line of matrix \mathbf{W} , and matrix \mathbf{W} is a set of the sparse codes on sample image, and max means the maximum value of the vector. Pooling process can influence the performance as shown in the curve of experimental section.

4.3. Constraint pooling

Different from max pooling and Abs pooling, we adopt a constraint to obtain the final representation. Enlightened by Hierarchical sparse coding, Yu *et al.* [19] introduce the second sparse coding with a weighted regularization of \mathbf{w}_i to get better performance on several benchmarks. We keep the spare representations information within regions and bring in a constraint

on the final sparse representation that obtaining the following optimization model:

$$\begin{aligned} \min_{\mathbf{z}} f(\mathbf{z}) &= \frac{1}{n} \sum_{i=1}^n \{\mathbf{w}_i^T \Sigma^{-1} \mathbf{w}_i\} \\ s.t. \quad &\|\mathbf{z}\|_2^2 = 1, \end{aligned} \quad (5)$$

where $\mathbf{z} \in R^p$, p is the number of codebook basis, \mathbf{w}_i is the sparse representation of image regions, $\Sigma = diag(\mathbf{z})^1$ is the diagonal matrix whose diagonal elements are the vector \mathbf{z} elements.

Using Lagrange method, put the constraint optimization problem into unconstraint problem:

$$\mathbf{z} = \arg \min_{\mathbf{z}} \frac{1}{n} \sum_{i=1}^n \{\mathbf{w}_i^T \Sigma^{-1} \mathbf{w}_i\} + \lambda_2 (\|\mathbf{z}\|_2^2 - 1), \quad (6)$$

where λ_2 is a lagrangian multiplier.

It can be transformed into general unconstraint optimization problem in form of matrix:

$$g(\mathbf{z}) = \text{tr}((diag(diag(\mathbf{W}\mathbf{W}^T)))\Sigma^{-1}) + \lambda_2 (\|\mathbf{z}\|_2^2 - 1), \quad (7)$$

where $\text{tr}()$ means the sum of the diagonal element of matrix. The gradient of $g(\mathbf{z})$ is

$$\nabla g(\mathbf{z}) = \begin{bmatrix} \frac{v_1}{z_1} \\ \vdots \\ \frac{v_n}{z_n} \end{bmatrix} + 2\lambda_2 \mathbf{z}, \quad (8)$$

where $\mathbf{v} = diag(\mathbf{W}\mathbf{W}^T)$. The solution of the $g(\mathbf{z})$ is

$$z_i = \sqrt{\frac{v_i}{\text{sum}(\mathbf{v})}}, \quad (9)$$

where $\text{sum}(\mathbf{v}) = \sum_{i=1}^n v_i$.

¹When the variable in the $diag(\cdot)$ is a vector, the result of the $diag(\cdot)$ is a diagonal matrix; or when the variable in the $diag(\cdot)$ is a matrix, the result of the $diag(\cdot)$ is a vector, the element of the vector is the diagonal element of matrix. Thus the $\Sigma = diag(\mathbf{z})$ is the diagonal matrix whose diagonal elements are the vector \mathbf{z} elements.

Note, optimization model can be reformulated in an expression:

$$\frac{1}{n} \sum_{i=1}^n \{\mathbf{w}_i^T \Sigma^{-1} \mathbf{w}_i\} = \text{tr}(C(\mathbf{W}) \Sigma^{-1}), \quad (10)$$

where

$$C(\mathbf{W}) = \frac{1}{n} \sum_{i=1}^n \mathbf{w}_i^T \mathbf{w}_i \quad (11)$$

is the covariance of sparse representation. The term involving Σ^{-1} implements a type of weighted regularization of \mathbf{w}_i . Similar to the energy constraint model in [24] that suppresses the unknown and undesired background signatures while enhancing the target signature. The result of the optimistic model could be considered as the two order statistic of sparse representations that keep all the information of the weighted regularization. The feature descriptors we choose are RGT descriptors based on different areas of annuli to compute the gradient histogram. It is reasonable to obtain the final sparse representation on this statistic form rather than to select the maximum one. What's more, the time complexity of solving the maximum is large when facing large scale of data. At the same time the constrained pooling and mean of absolute values are also compared in the experimental section.

5. Airplane detection based on classifier

A simple linear support vector machine (SVM), which is suited to classify sparse representation for better performance, is present in the paper. We detect the airplane by using binary classifier. Thus, given the training data $\{(\mathbf{z}_i, y_i)\}_{i=1}^m, y_i \in \{-1, 1\}$, where \mathbf{z}_i means the i -th final sparse representation of sample image, n is the number of sample image, y_i is the input label belong to -1 and 1 indicating the non-airplane and airplane. The form of classifier is following:

$$y(\mathbf{z}) = \text{sign} \left[\sum_{k=1}^m a_k y_k \Psi(\mathbf{z}, \mathbf{z}_k) + b \right], \quad (12)$$

where a_k is a positive real constant and b is a real constant. We choose the function Ψ as linear function, $\Psi(\mathbf{z}, \mathbf{z}_k) = \mathbf{z}_k^T \mathbf{z}$.

When using the nonlinear SVM to classify the targets, the complexity is $O(n^2 \sim n^3)$ in training and $O(n)$ in testing, implying that it is trouble to deal with large-scale data with more than thousands of training and test images.

In our paper we use liner SVM classifier to do experiments, which has better performance and is high-efficiency. **Algorithm 1** shows the detection process based on sparse coding and a linear SVM classifier.

Algorithm 1 The process of detection via sparse coding in remote sensing images.

Step 1: Train the dictionary \mathbf{B} by using RGT descriptors $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n]$ from target or non-target target samples, and iteratively train with sparse representation $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n]$:

$$\mathbf{B} \leftarrow \arg \min_{\mathbf{B}} \|\mathbf{X} - \mathbf{B}\mathbf{W}\|_F^2 + \mu(\|\mathbf{B}\|_1 - 1),$$

where μ is a lagrangian multiplier vector.

Step 2: Given a remote sensing images, locate the roughly location of the airplanes by Circle-frequency filter [9] and use sliding window on candidate region to obtain a set of center P .

Step 3: Select a point $p \in P$, and sample the size of 40×40 pixels region from the image at the point of p .

Step 4: Compute the RGT descriptor of the sample image, and then the sparse representation of descriptors $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_n]$ is obtained by:

$$\mathbf{w}_i \leftarrow \arg \min_{\mathbf{w}} \|\mathbf{x}_i - \mathbf{B}\mathbf{w}\|^2 + \lambda \|\mathbf{w}\|_1,$$

and then pool the sparse representation of RGT descriptors.

Step 5: The final feature of sparse representation is obtained by pooling:

$$z_i = \sqrt{\frac{v_i}{sum(\mathbf{v})}},$$

where $\mathbf{v} = diag(\mathbf{W}\mathbf{W}^T)$.

Step 6: Use a linear SVM classifier to classify the obtained sample.

Step 7: Back to step 3, until the set of P is empty.

6. Experiments

We verify the performance of the proposed method on samples of remote sensing images. In the task, we report the prediction accuracies for our model

with sparse coding. We also compare our rotation-invariant sparse feature with other features under the same experiment setting.

6.1. Datasets

We test our detector on data set containing 54 images of airports from Google Maps. These images rang from 800×800 pixels to 1200×1200 pixels with one meter resolution. The airplanes in the images have unpredictable directions. We select three kinds of airplane directions, which are 0° , 45° and 90° directions. Together with their left-right reflections and up-down reflections, positive samples set have about 3666 airplanes with eight directions. Considering the general size of the airplane in one resolution images, we choose 40×40 pixels as the sample size. In non-airplane regions of images, we randomly select 25508 samples as a negative training set.

We plot receiver operating characteristics (ROC) curve [25] to quantify feature performance, i.e. $\text{TPR} = \frac{\text{TruePositive}}{\text{Positive}}$ and $\text{FPR} = \frac{\text{FalseNegative}}{\text{Negative}}$, where TruePositive (TP) and Positive (P) mean the number of detected true airplanes and the number of airplanes set, respectively; FalseNegative (FN) and Negative (N) mean the number of the non-airplanes that detected as airplane and the number of the non-airplanes set. They present the same information as Detection Error Tradeoff (DET) [16]. We perform 5-fold cross validation and report average results across all folds [26]. The better performance, the higher true positive rate and lower false positive rate. We use default $\text{accuracy} = \frac{\text{TP} + \text{TN}}{\text{P} + \text{N}}$ [25] as a reference for the performance, where TrueNegative (TN) is the number of the correctly detected non-airplanes.

6.2. Analysis of results

To obtain the sparse representation, we train the dictionaries that well adapt to the training set. We use a single but unusual descriptor. This descriptor is based on annuli to compute the gradient histogram by radial gradient transform. Unlike the patch extracted from tradition method, our patch is the annulus around the center of the samples. We set four pixels as the width of the annuli to obtain 8 annuli in all within each sample, and each annulus corresponds to a RGT descriptor. The dimension of the RGT descriptors is 72. These descriptors are pre-normalized to be unit vectors before sparse coding. The sparse regularization λ is set to 0.15 empirically. Then we train the codebook with 1024 bases based on these RGT descriptors.

Each sample is divided into eight patch annuli, where each path annulus corresponds to a sparse vector. These sparse vectors are pooled together to

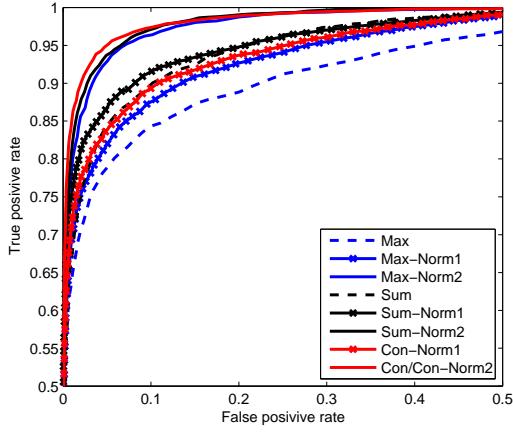


Figure 4: The performance of ROC curve based on sparse coding with different pooling methods and normalization.

get the final sparse representation. Specifically, three pooling methods are used: sum pooling, max pooling and constraint pooling. These final sparse vectors can be normalized by L_1 normalization or L_2 normalization. Notice that, the sparse vector with constraint pooling is equal to sparse vector with L_2 normalization, because the L_2 normalization of the sparse vector \mathbf{z} is one in the constraint term of the optimization model. Figure 4 shows the result of the performance with different pooling methods and normalization. Both of the L_1 normalization and L_2 normalization measured by ROC curve outperform than no normalization. The ROC performance with constraint pooling has the best result with high true positive rate and low false positive rate.

As shown in Figure 5, we compare different methods of feature extraction. These features have rotation-invariant property: Rotation-invariant fast feature (RIFF) [8]², Approximate radial gradient transform (ARGT) [8], and local binary pattern fourier feature (LBP-HF) [27]. Besides these methods, we also use the HOG feature [16] to demonstrate that the HOG feature could not well deal with the rotation samples, which has the worst accuracy result. The ROC curves of RIFF [8] and ARGT [8] are almost close to each other, and the accuracy of RIFF and ARGT is 96.50% and 96.45%

²The RGT descriptor is based on RIFF

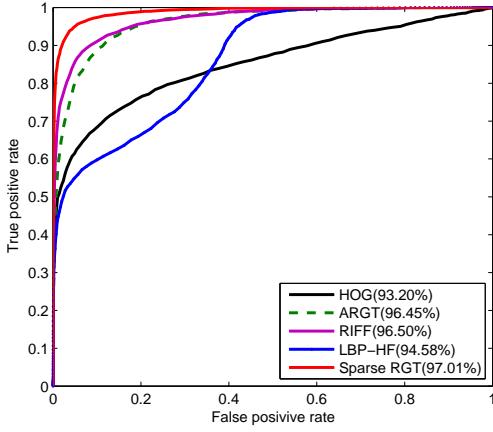


Figure 5: The ROC curve with different features based on False positive rate and true positive rate.

respectively. However, LBP-HF, based on gray information, has poor ROC curve performance comparing with RIFF and ARGT. By using RGT descriptors, sparse representation of airplane achieves 97.01% comparing with other feature methods, which achieves the best performance.

Table 1: Detection results in remote sensing images

	PCA[13]	Symmetry[14]	Shape[12]	RGT	Sparse coding
Detection rate	85%	80.3%	89.3%	92.97%	94.08%

Figure 6 shows the result of the detection on the whole remote sensing images. In each image, there are several airplanes that locate with arbitrary directions but are well detected by using sparse representation feature. Before detection, we need to preprocess the image to decrease the detection time by using circle frequency filter [9], which sets the threshold to 0.05, and gaussian filter. Circle frequency filter roughly locates the airplane based on the shape information, and gaussian filter smoothes the candidate regions. They greatly decrease the detection time. In the detection process, we use the sliding window on the candidate regions to extract features. Before obtaining the sparse features, we have trained the dictionary by using the sample set based on RGT descriptors. We use linear SVM classifier to detect the airplane using sparse representation features. The detection rate is shown in Table 1.

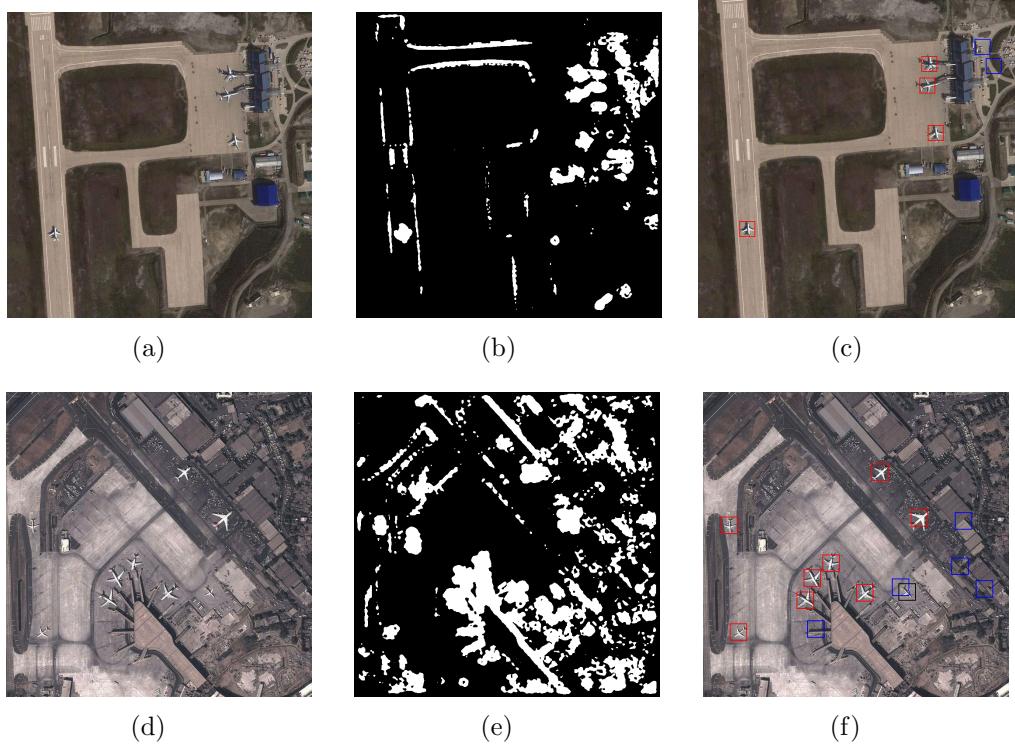


Figure 6: The result of airplane detection. (a) and (d) show the remote sensing images; (b) and (e) are the candidate regions result preprocess by the circle frequency filter and gaussian filter; (c) and (f) are the result of detection: the red box means the right detection, the blue indicate the false detection, and the black is the missing airplane.

The detection method based on RGT features has the better detection rate compared with methods, such as PCA and model matching[13], symmetry-based algorithm[14] and coarse-to-fine shape prior[12]. Spare representation feature achieves the best performance of 94.08%, outperforming the RGT feature that achieves 92.97%, and other methods range between 80% – 89%.

7. Conclusion

This paper presents a new feature representation of the airplane in remote sensing images based on sparse coding for airplane detection. We apply the radial gradient transform to the feature extraction process, thus the obtained feature descriptors have the rotation-invariant property. To get the better representation of airplane, we adopt sparse coding combined with constraint

pooling to optimize a linear combination of basis for obtaining the sparse representation. These bases are learned from the RGT descriptor such that the obtained final sparse representation possesses rotation-invariant property. We also analyze the pooling methods based on max pooling, mean pooling, and constraint pooling. The constraint pooling captures the statistic information of sparse vectors that well represent the airplane features. The experimental results show that combining with constraint pooling the sparse representation has better ROC curves and higher detection rate, and the rotation-invariant sparse coding provides a promising way on general object detection in remote sensing images.

8. Acknowledgments

The work was supported by the National Natural Science Foundation of China under the Grants 61273245 and 91120301, the 973 Program under the Grant 2010CB327904, and Program for New Century Excellent Talents in University of Ministry of Education of China under the Grant NCET-11-0775. The work was also supported by Beijing Key Laboratory of Digital Media, Beihang University, Beijing 100191, P.R. China.

- [1] A. Filippidis, L. C. Jain, N. Martin, Fusion of intelligent agents for the detection of aircraft in sar images, *IEEE Transactions on, Pattern Analysis and Machine Intelligence* 22 (4) (2000) 378–384.
- [2] B. Chalmond, B. Francesconi, S. Herbin, Using hidden scale for salient object detection, *IEEE Transactions on, Image Processing* 15 (9) (2006) 2644–2656.
- [3] W. Li, S. Xiang, H. Wang, C. Pan, Robust airplane detection in satellite images, in: *Proceedings of Image Processing (ICIP)*, 2011.
- [4] D. L. Donoho, For most large underdetermined systems of linear equations the minimal l1-norm solution is also the sparsest solution, *Communications on pure and applied mathematics* 59 (6) (2006) 797–829.
- [5] S. Lazebnik, C. Schmid, J. Ponce, Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories, in: *Proceedings of Computer Vision and Pattern Recognition (CVPR)*, 2006, Vol. 2, pp. 2169–2178.

- [6] J. Yang, K. Yu, Y. Gong, T. Huang, Linear spatial pyramid matching using sparse coding for image classification, in: Proceedings of Computer Vision and Pattern Recognition (CVPR), 2009, pp. 1794–1801.
- [7] D. G. Lowe, Distinctive image features from scale-invariant keypoints, International journal of computer vision 60 (2) (2004) 91–110.
- [8] G. Takacs, V. Chandrasekhar, S. Tsai, D. Chen, R. Grzeszczuk, B. Girod, Unified real-time tracking and recognition with rotation-invariant fast features, in: Proceedings of Computer Vision and Pattern Recognition (CVPR), 2010, pp. 934–941.
- [9] H. Cai, Y. Su, Airplane detection in remote sensing image with a circle-frequency filter, in: Proceedings of Space information Technology, International Society for Optics and Photonics, 2005.
- [10] Y. Li, X. Sun, H. Wang, H. Sun, X. Li, Automatic target detection in high-resolution remote sensing images using a contour-based spatial model, IEEE, Geoscience and Remote Sensing Letters 9 (5) (2012) 886–890.
- [11] C. Xu, H. Duan, Artificial bee colony (abc) optimized edge potential function (epf) approach to target recognition for low-altitude aircraft, Pattern Recognition Letters 31 (13) (2010) 1759–1772.
- [12] G. Liu, X. Sun, K. Fu, H. Wang, Aircraft recognition in high-resolution satellite images using coarse-to-fine shape prior.
- [13] D. SHAO, Y. ZHANG, W. WEI, An aircraft recognition method based on principal component analysis and image model matching, Chinese Journal of Stereology and Image Analysis 3 (2009) 7.
- [14] J.-W. Hsieh, J.-M. Chen, C.-H. Chuang, K.-C. Fan, Aircraft type recognition in satellite images, in: Proceedings of Computer Vision, Image and Signal Processing, 2005, Vol. 152, pp. 307–315.
- [15] H. Bay, T. Tuytelaars, L. Van Gool, Surf: Speeded up robust features, in: Proceedings of European Conference on Computer Vision (ECCV), 2006, pp. 404–417.

- [16] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: Proceedings of Computer Vision and Pattern Recognition, 2005, Vol. 1, pp. 886–893.
- [17] J. Yang, K. Yu, T. Huang, Supervised translation-invariant sparse coding, in: Proceedings of Computer Vision and Pattern Recognition (CVPR), 2010, pp. 3517–3524.
- [18] H. Lee, A. Battle, R. Raina, A. Ng, Efficient sparse coding algorithms, Advances in neural information processing systems 19 (2007) 801.
- [19] K. Yu, Y. Lin, J. Lafferty, Learning image representations from the pixel level via hierarchical sparse coding, in: Proceedings of Computer Vision and Pattern Recognition (CVPR), 2011, pp. 1713–1720.
- [20] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE 86 (11) (1998) 2278–2324.
- [21] Y. Jia, C. Huang, T. Darrell, Beyond spatial pyramids: Receptive field learning for pooled image features, in: Proceedings of Computer Vision and Pattern Recognition (CVPR), 2012, 2012, pp. 3370–3377.
- [22] Y. Boureau, N. Le Roux, F. Bach, J. Ponce, Y. LeCun, Ask the locals: multi-way local pooling for image recognition, in: Proceedings of International Conference Computer Vision (CVPR), 2011, pp. 2651–2658.
- [23] Y. Boureau, F. Bach, Y. LeCun, J. Ponce, Learning mid-level features for recognition, in: Proceedings of Computer Vision and Pattern Recognition (CVPR), 2010, pp. 2559–2566.
- [24] W. H. Farrand, J. C. Harsanyi, Mapping the distribution of mine tailings in the coeur d'alene river valley, idaho, through the use of a constrained energy minimization technique, Remote Sensing of Environment 59 (1) (1997) 64–76.
- [25] T. Fawcett, An introduction to roc analysis, Pattern recognition letters 27 (8) (2006) 861–874.
- [26] U. Schmidt, S. Roth, Learning rotation-aware features: From invariant priors to equivariant descriptors, in: Proceedings of Computer Vision and Pattern Recognition (CVPR), 2012, pp. 2050–2057.

- [27] T. Ahonen, J. Matas, C. He, M. Pietikäinen, Rotation invariant image description with local binary pattern histogram fourier features, in: Image Analysis, Springer, 2009, pp. 61–70.