

## Rotation-invariant infrared aerial target identification based on SRC

JIN Lu<sup>1,2,3</sup>, LI Fan-Ming<sup>1,3\*</sup>, LIU Shi-Jian<sup>1,3</sup>, WANG Xiao<sup>1,2,3</sup>

- (1. Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai 200083, China;  
2. University of Chinese Academy of Sciences, Beijing 100049, China;  
3. CAS Key Laboratory of Infrared System Detection and Imaging Technology, Shanghai Institute of Technical Physics, Shanghai 200083, China)

**Abstract:** Aircraft identification is implemented on thermal images acquired from ground-to-air infrared cameras. SRC is proved to be an effective image classifier robust to noise, which is quite suitable for thermal image tasks. However, rotation invariance is challenging requirements in this task. To solve this issue, a method is proposed to compute the target main orientation firstly, then rotate the target to a reference direction. Secondly, an over-complete dictionary is learned from histogram of oriented gradient features of these rotated targets. Thirdly, a sparse representation model is introduced and the identification problem is converted to a  $l_1$ -minimization problem. Finally, different aircraft types are predicted based on an evaluation index, which is called residual error. To validate the aircraft identification method, a recorded infrared aircraft dataset is implemented in an airfield. Experimental results show that the proposed method achieves 98.3% accuracy, and recovers the identity beyond 80% accuracy even when the test images are corrupted at 50%.

**Key words:** infrared image, aircraft identification, rotation invariant, sparse representation classification

**PACS:** 42.30.Sy, 07.05.Mh

## 基于稀疏表示的红外空中目标分类算法

金 璐<sup>1,2,3</sup>, 李范鸣<sup>1,3\*</sup>, 刘士建<sup>1,3</sup>, 王 霄<sup>1,2,3</sup>

- (1. 中国科学院上海技术物理研究所, 上海 200083;  
2. 中国科学院大学, 北京 100049;  
3. 中国科学院红外探测与成像技术重点实验室, 上海 200083)

**摘要:** 针对红外空中目标, 提出了一种基于稀疏表示的快速分类算法. 该工作的技术难点表现在训练样本较少, 算法需要具有旋转不变性、较高的抗噪性和实时性. 针对这些难点, 首先根据红外空中目标的梯度信息和统计特性, 计算出图像主方向, 然后将主方向旋转至同一参考方向. 接着基于稀疏表示原理, 把分类问题转化为  $l_1$  范数最小化问题, 最后用快速收敛方法得到分类结果. 实验结果表明该方法能够达到 98.3% 的正确率, 给测试图像 50% 的像素叠加噪声后, 分类正确率仍大于 80%.

**关键词:** 红外图像; 空中目标; 旋转不变性; 稀疏表示分类

中图分类号: TP391.4 文献标识码: A

## Introduction

Infrared target recognition and classification are significant parts in video surveillance and aeronautics appli-

cations. In aeronautics applications, aircraft are the main targets to surveil. Especially in ground-to-air applications, a system which has good performance at anti-jamming, fast identification friend or foe and stable track-

**Received date:** 2019-01-09, **revised date:** 2019-07-08

**收稿日期:** 2019-01-09, **修回日期:** 2019-07-08

**Foundation items:** Supported by the Thirteen Five National Defense Research Foundation (Jzx2016-0404/Y72-2); Shanghai Key Laboratory of Criminal Scene Evidence funded Foundation (2017xcwzk08).

**Biography:** JIN Lu (1991-), female, Wuhan, PhD. Research area involves computer vision and pattern recognition. E-mail: jinlu0716@163.com.

\* **Corresponding author:** E-mail: lfmjws@163.com

ing capability is extremely required. Comparing with visible light cameras, which are restricted by the necessity of clear meteorological conditions, infrared cameras show superiority of robustness to illumination and weather conditions. However, in infrared aerial identification task, particularly in ground-to-air applications, targets generally occupy several pixels in imaging device and has not that much information of figures. Besides, clouds occlusion and large pose variation also increase the difficulty of identification. Due to these reasons, we must extract as much information as possible from finite data.

According to the principle of target identification, conventional algorithms are usually divided into three steps: Firstly, find the regions of interest in image sequences. Then extract their features and finally predict different types of these targets using specific classifiers. In our previous work, the targets are already detected, so our concentration is to identify to which of predefined aircraft types the target belongs.

In feature extraction field, plenty of creative methods have been proposed, which are based on either manual design (e. g. , SIFT<sup>[1]</sup>, SURF<sup>[2]</sup>, HOG<sup>[3]</sup>) or learning (e. g. , bag-of-words<sup>[4]</sup>, neural networks<sup>[5]</sup>). Among these approaches, learning-based methods require sufficient labeled data. This is tough for our work because the cost of IR aircraft images is very expensive, especially for jets. As for SIFT and SURF methods, they focus on the description of interest points, so these two descriptors are more suited for tasks which need to check matching degree between key points, such as image matching and image retrieval. HOG feature is widely used in object recognition and classification and proves to be very robust in related works. However, target rotation and pose variation are instrumental in ground-to-air IR images while HOG feature is clearly not rotation-invariant. To handle this problem, Takacs et al<sup>[6]</sup> proposed a rotation-invariant descriptor which introduces the radial gradient transform in polar coordinates. Some similar configurations are proposed in recent works<sup>[7-9]</sup>. Nevertheless, These HOG descriptors in polar coordinates reject information of local image regions and target direction. Another solution to target rotation is data augmentation, which means rotating training samples to different angles in learning process. However, this will lead to high computational complexity and not meet the requirement to real-time tasks. In contrary to these methodologies, we address the rotation invariant issue by incorporating the concept of main orientation into HOG descriptor. The detail is to be presented in section II.

In classification methods, existing aircraft classification algorithms mainly based on the nearest feature, support vector machine (SVM) or neural networks. Among them, methods based on neural networks are the research focus in recent years, plenty of architectures based on deep convolutional neural networks are proposed, and they achieve outstanding performance. However, these architectures are trained on large number of images with refined annotations, which is quite costly for

us as mentioned before. Sparse Representation Classification (SRC) seeks a sparse coefficient of an equation in which the image is represented by this coefficient according to an overcomplete dictionary, then performs classification process by checking which class outputs the least reconstruction error. Therefore, SRC has the advantage of both neural networks and nearest feature classifier. In the work of Wright et al<sup>[10]</sup>, SRC is proved to remain 100% recognition rate even when the image is corrupted by 60%. This prominent performance is quite suitable for IR aircraft identification due to the serious noise in IR images. For now, SRC is mostly used in face recognition<sup>[11-13]</sup>. It has never been applied in aircraft identification, so we were interested in how it performs in predicting aircraft types. The nearest feature and SVM are typical small sample learning algorithms and have excellent classification performance, so we also test them in our work as a contrast.

The remainder of this article is organized as follows. In section II, we propose the rotation-invariant HOG features based on main orientation. Then we present a brief introduction of Sparse Representation Classification and dictionary learning. Afterwards, we illustrate the performance of our algorithm in experiments. In conclusion and future section, we make a summary of our work and give the suggestions for future work.

## 1 Proposed Method

Figure 1 illustrates the flowchart of our identification method. It consists of two stages: dictionary construction and target identification. In dictionary construction section, we firstly compute the main orientations of training samples depending on the gradient information, then rotate these samples to the reference direction. Afterwards, we extract HOG features from the rotated samples to construct the initial dictionary. To improve the classification ability of this dictionary, we incorporate FDDL, a dictionary learning method into dictionary construction. In target identification section, the target is also rotated to the reference direction based on its main orientation. Then we extract its HOG feature and compute its sparse representation coefficients from the dictionary constructed ahead. At last this target is identified depending on its smallest reconstruction error.

### 1.1 Main Orientation Extraction

There is a specific character of IR aircraft targets: aeroengine shows the strongest thermal radiation. Based on this character, we define the main orientation of a target is largely based on its aeroengine. Detailed process of main orientation extraction is as follows.

**Step 1:** Gradient magnitude and orientation computation

Consider a pixel located at position  $(x, y)$  where  $x$  indicates the row position and  $y$  indicates the column position, let  $I(x, y)$  denotes the intensity value of pixel located at  $(x, y)$ . The gradient magnitude  $M$  and gradient orientation  $\theta$  of each pixel are calculated as the formulas below

$$|M(x, y)| = \sqrt{G_h^2(x, y) + G_v^2(x, y)} \quad , \quad (1)$$

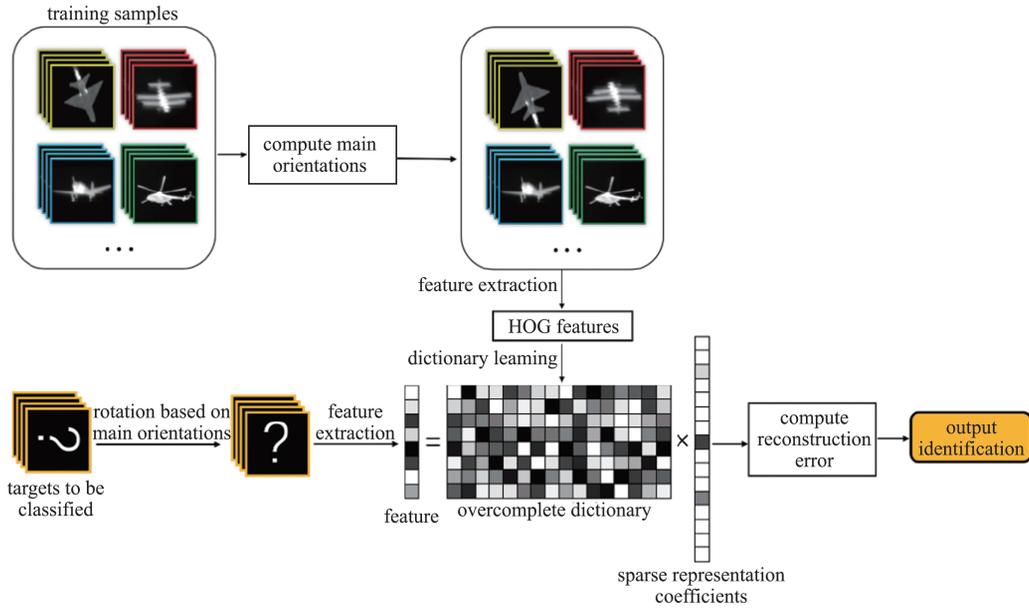


Fig. 1 Framework of the proposed identification method  
图1 红外空中目标的分类算法流程

$$\theta(x,y) = \tan^{-1}\left(\frac{G_v(x,y)}{G_h(x,y)}\right), \quad (2)$$

where  $G_h$  represents the gradient values in horizontal direction and  $G_v$  represents the gradient values in vertical direction.  $G_h$  and  $G_v$  are defined as follows

$$G_h(x,y) = I(x+1,y) - I(x-1,y), \quad (3)$$

$$G_v(x,y) = I(x,y+1) - I(x,y-1). \quad (4)$$

### Step 2: Gradient vote,

Use the gradient orientation of an image to weighted vote into  $n$  corresponding orientation bins equally spaced between  $0^\circ$  and  $360^\circ$ , the vote is weighted by the intensity value of pixels. Then midpoint of the highest orientation bin is the main orientation of this target.

For instance, let us set  $n$  to 12, as is shown in Fig. 2, main orientation of target (a) is  $195^\circ$ , which is the midpoint of  $180^\circ$  and  $210^\circ$ . In the same operation, main orientation of target (b) is  $135^\circ$  and main orientation of target (c) is  $45^\circ$ . On the basis of Eq. (2) (3) and (4), positive direction is clockwise direction. In Fig. 2, the direction of green arrow is the reference direction (3 o'clock direction), and direction of orange arrow is the main orientation of targets. After the targets are rotated in anti-clockwise direction according to their main orientation, they would be in almost the same direction. However, it is notable that there might be a subtle difference between rotated targets, just like (b) and (c) in Fig. 2. Although HOG features have invariant descrip-

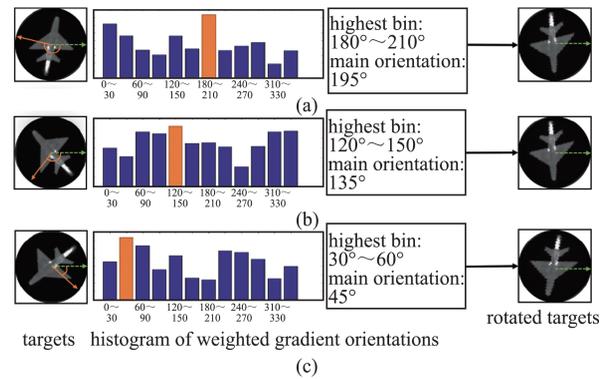


Fig. 2 Rotation according to main orientation. Green arrow: reference direction. Orange arrow: main orientation of this target  
图2 根据主方向对目标进行旋转. 绿色虚线箭头: 参考方向. 橙色实线箭头: 目标主方向

tion in a small angle rotation, we want to know how small the difference is acceptable in our identification task, so a series of experiments is conducted. The results are showed in Table 1.

In this experiment, we manually rotate all the test images to different angles varying from  $3^\circ$  to  $24^\circ$ , correspondingly, number of orientation bins is varied from 120 to 15. As is shown in Table 1, when the rotation angle is within  $12^\circ$ , identification accuracy remains stable beyond 90%. Therefore, 30 is the suitable number of rotation bins we choose to do gradient weighted vote.

Table 1 identification accuracy according to different rotation degrees

表1 不同旋转角度下的准确率对比

rotation degree	$3^\circ$	$5^\circ$	$7.5^\circ$	$10^\circ$	$12^\circ$	$15^\circ$	$20^\circ$	$24^\circ$
number of rotations	120	72	48	36	30	24	18	15
accuracy	95.7%	95.1%	93.8%	92.8%	91.4%	88.1%	77.4%	62.8%

## 1.2 Histograms of Oriented Gradients

The main idea behind HOG is that any shape or local object in an image can be well discriminated by knowledge of only edge direction and without knowing their actual position<sup>[14]</sup>. Process of HOG feature extraction is shown in Fig. 3. First, we compute the gradient magnitude and orientation. Then we divide the image into cells, use the gradient orientation  $\theta$  to vote into 9 corresponding orientation bins equally spaced between  $0^\circ$  and  $180^\circ$ , the vote is weighted by gradient magnitude  $M$ . To enhance illumination invariant ability, we normalize all the histograms which are calculated over cells in a block with an overlapping of 50%. At last the HOG feature descriptor of the target is constructed by linking all HOG features of blocks together. This is the final eigen-vector for classification process.

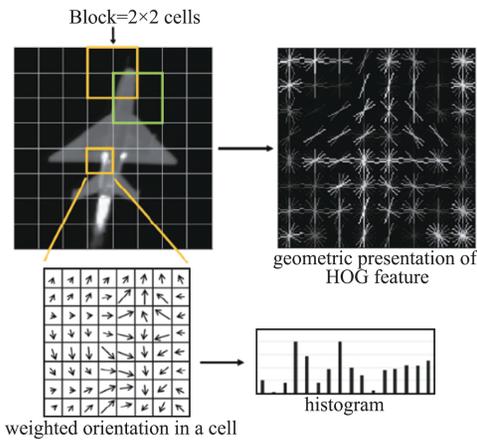


Fig. 3 HOG extraction from an IR aerial target  
图3 红外空中目标的HOG特征提取示意图

## 1.3 Sparse Representation-based classification

Sparse representation is successfully used for face recognition and fingerprint classification mainly because the sparsest representation is naturally discriminative: among all subsets of base vectors, it selects the subset which most compactly expresses the input signal and rejects all other possible but less compact representations<sup>[10]</sup>. Besides, sparse representation not only can find the inner information in just a small amount of model data but also performs robustness to occlusion or corruption. The conventional framework for Sparse Representation Classification can be divided into three steps: dictionary construction, sparse representation and identity prediction. In this article we incorporate dictionary learning method into SRC process to improve the classification ability of the dictionary. The process is as follows.

### Step 1: dictionary initialization

Suppose we have  $c$  different classes and each class contains  $m$  training samples. Feature vector  $f \in \mathbb{R}^d$  represents a training image and  $A_k = [f_{k1}, f_{k2}, \dots, f_{km}] \in \mathbb{R}^{d \times m}$ , ( $k = 1, 2, \dots, c$ ) is the matrix of training images from the  $k$ th class. In other words,  $A_k$  represents a sub-dictionary for class  $k$ . Then we define a new matrix of dictionary  $A \in \mathbb{R}^{d \times n}$  as the concatenation of sub-dictionaries from all

the classes

$$A = [A_1, A_2, \dots, A_c] = [f_{11}, f_{12}, \dots, f_{cm}], A \in \mathbb{R}^{d \times n}, \quad (5)$$

$A$  is the initial dictionary for the next step.

### Step 2: dictionary learning

Dictionary learning is to learn from the training samples so that given signals could be well represented using the optimized dictionary. Many dictionary learning methods have been proposed in the past few years such as MOD<sup>[15]</sup>, K-SVD<sup>[16]</sup> and FDDL<sup>[17]</sup>. However, MOD and K-SVD are not suitable for classification tasks because they only require that the learned dictionary can well represent training samples, ignoring their classification ability. Yang<sup>[18]</sup> proposed a dictionary learning framework called FDDL which uses the Fisher discrimination criterion to get an optimized dictionary. In this algorithm, sparse coding coefficients have small within-class scatter but large between-class scatter. Meanwhile, each sub-dictionary for class  $k$  is able to well represent the training samples from the corresponding class. In contrast, they have poor ability to represent other classes. Therefore, we use FDDL as the learning method to optimize our dictionary. The objective function of FDDL model is

$$J_{(D,X)} = \arg \min_{(D,X)} \left\{ \sum_{i=1}^c r(A_i, D, X_i) + \lambda_1 \|X\|_1 + \lambda_2 f(X) \right\}, \quad (6)$$

where  $r(A_i, D, X_i)$  is the discriminating fidelity term,  $\lambda_1 \|X\|_1$  is the sparsity constraint term, and  $\lambda_2 f(X)$  is the discriminating coefficient term. Expanded form of these three terms can be found in Ref. [18]. This equation is not convex to  $(D, X)$  at the same time. However, when  $D$  is fixed, it is convex to  $X$  and vice versa. The detailed optimization of Eq. (6) is as follows.

Firstly, initialize the dictionary utilizing training data. Secondly, fix  $D$  and update the sparse coding coefficients  $X$  by solving the equation below.

$$J_{(X)} = \arg \min_{(X)} \left\{ r(A_i, D, X_i) + \lambda_1 \|X_i\|_1 + \lambda_2 f_i(X_i) \right\}, \quad (7)$$

$$i = 1, 2, \dots, c$$

Thirdly, fix  $X$  and update  $D$  by solving the equation below.

$$J_{(D)} = \arg \min_{(D)} \left\{ \left\| A - D_i X^i - \sum_{j=1, j \neq i}^c D_j X^j \right\|_F^2 + \left\| A_i - D_i X^i \right\|_F^2 + \sum_{j=1, j \neq i}^c \left\| D_j X^j \right\|_F^2 \right\}, i = 1, 2, \dots, c \quad (8)$$

Then return to the second step until reaching stop criterion.

### Step 3: sparse representation

Suppose  $y \in \mathbb{R}^d$  is a feature vector of the target we are trying to identify, and  $D \in \mathbb{R}^{d \times n}$  is the optimized dictionary acquired from Step 2. Then  $y$  can be represented as a linear combination of atoms in  $D$  with coefficients  $x$ .

$$\mathbf{y} = \sum_{i=1}^c \sum_{j=1}^m \mathbf{x}_{ij} f_{ij} \quad (9)$$

This equation can be more compact as

$$\mathbf{y} = \mathbf{D}\mathbf{x} \quad (10)$$

where

$$\mathbf{x} = [\mathbf{x}_{11}, \mathbf{x}_{12}, \dots, \mathbf{x}_{cm}]^T \in \mathbb{R}^n \quad (11)$$

And  $\cdot^T$  denotes the transposition operation. Suppose the target to be identified belongs to class  $k$  in reality and each class contains enough training samples, then  $\mathbf{y}$  will be more relevant to the atoms of  $\mathbf{D}_k$  than other sub-dictionaries. Videlicet, the values of coefficients  $\mathbf{x}$  that are irrelevant to class  $k$  is almost to zero in (8) and  $\mathbf{x}$  is a very sparse solution to equation (7). However, with an overcomplete dictionary, equation (7) has infinite solutions, among which we have to find the sparsest one. Here we use  $l_1$ -norm minimization to address this issue.

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \left\{ \|\mathbf{y} - \mathbf{D}\mathbf{x}\|_2^2 + \gamma \|\mathbf{x}\|_1 \right\} \quad (12)$$

where  $\gamma$  is scalar constant.

The most well-known algorithms of  $l_1$ -norm minimization are orthogonal matching pursuit (OMP) and least angle regression (LARS), which suffer from either too much computational overhead or deficient estimation accuracy in large scale applications. New algorithms proposed in recent years are gradient projection<sup>[19]</sup>, homotopy<sup>[20]</sup>, iterative shrinkage-thresholding<sup>[21]</sup>, proximal gradient<sup>[22]</sup> and alternating direction<sup>[23]</sup>. Among all these algorithms, OMP is the most widely used algorithm, and homotopy is the fastest algorithm not only appropriate to large scale applications but also capable to arrive at the sparsest solutions.

#### Step 4: classification principle

At last, we use the sparse representation of the target's feature vector to reconstruct with each class of sub-dictionaries. The target to identified is predicted to belong to the class with the least reconstruction error. The reconstruction of class  $k$  is defined as follows: keep the coefficients corresponding to class  $k$  while setting the remaining coefficients to zero. Here we introduce a function  $\chi_k, \mathbf{X}_k$  has the value of  $\mathbf{x}$  at locations corresponding to class  $k$  and value zero for others. Reconstruction error of class  $k$  is defined as

$$error_k(\mathbf{y}) = \|\mathbf{y} - \mathbf{D}\mathbf{X}_k(\mathbf{x})\|_2 \quad (13)$$

Then the identity of the target is predicted as

$$identity(\mathbf{y}) = \arg \min_k \{error_k\} \quad (14)$$

## 2 Experiments

### 2.1 Dataset and experimental setup

In our experiments, aircraft images are acquired

from ground-to-air IR videos in airfields, which consist of helicopter-, airliner-, transport-, trainer- and two types of jets. Depending on the different position of these aircraft (front view and side view), we divide all the images into 8 categories: helicopter-type, transport-type, front view of airliner-type, side view of airliner-type, front view of trainer-type, side view of trainer-type, jet-type1 and jet-type2, as is shown in Fig. 4. Each category contains the number of images varies from 300 to 537. The detailed number is shown in Table 2. From each category, we randomly choose 60 images to constitute the initial dictionary and 200 images from the rest as test images in each experiment. Besides, we rotate all the test images into 30 orientations at even 12 degrees.

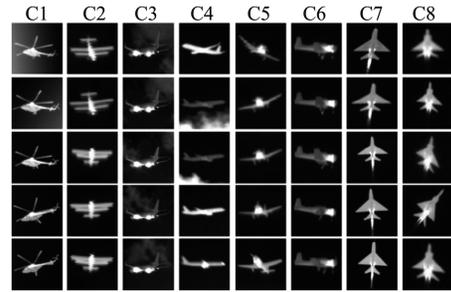


Fig. 4 Sample images for the eight classes  
图4 八种不同类别的红外空中目标示例

All experiments are performed on the hardware platform of Intel(R) Core(TM) i7-6700HQ CPU@2.60GHz with 8GB of DDR RAM, and the software platform is matlab R2016a.

### 2.2 Feature extraction and Dictionary learning

At the very beginning of dictionary learning procedure, all the training images are resized to 40\*40 pixels. Main orientations of training samples are computed afterwards. 0° is chosen to be the reference direction, then each of the training samples is rotated until its main orientation coincides with the reference direction. After that, we extract HOG features of these training samples. Here we set the size of cell to 10\*10 pixels and the size of block to 2\*2 cells with an overlapping of 50%. Each cell contains 9-bin histogram of oriented gradients (0~180°, 20° step size) and each block contains a concatenated vector of all 2\*2 cells. In this case, HOG feature of each training sample is a 324-dimension vector. From each class, we randomly choose 60 images to constitute the initial dictionary, which means the initial dictionary is a 324\*480 matrix, as is shown in Fig. 5.

According to section 2.3, optimization of FDDL can be divided into two alternating procedures: updating coefficients  $X$  by fixing dictionary  $D$ ; and updating dictionary  $D$  by fixing coefficients  $X$ . Here we choose the param-

Table 2 Specification of experimental sources

表2 实验数据说明

aerial type	helicopter-	transport-	airliner-	airliner-	trainer-	trainer-	jet-	jet-
			(front)	(side)	(front)	(side)	type1	type2
number of images	537	300	300	484	300	300	484	516

0.0127	0.0360	0.0213	0.0577	0.0232
0.0209	0.0320	0.0183	0.0311	0.0205
0.0361	0.1305	0.0451	0.1512	0.0601 ...
0.1198	0.0398	0.0464	0.0405	0.0898
0.2999	0.0285	0.0553	0.2604	0.1491
		...		

initial dictionary size-324\*480

Fig. 5 Initial dictionary matrix  
图5 初始字典矩阵

eters of Eq. (6)  $\lambda_1=0.01$  and  $\lambda_2=0.01$ . Convergence of Eq. (7) and Eq. (8) is illustrated in Fig. 6.

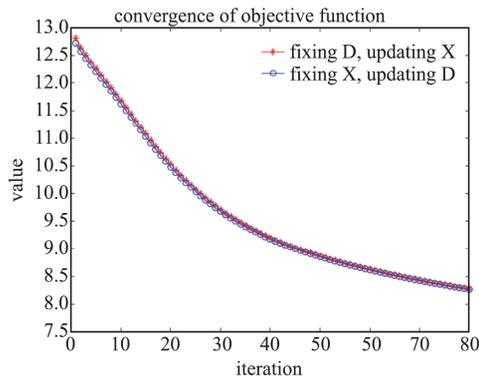


Fig. 6 Convergence of FDDL model  
图6 FDDL模型的收敛曲线

### 2.3 Experimental results for classification

To prove the rotation invariance of our algorithm, we manually rotate all the test images into 30 orientations varying from  $0^\circ$  to  $348^\circ$  at even angles of  $12^\circ$ . In reality, only transport-type, jet-type1 and jet-type2 may quite possibly appear in different orientations in ground-to-air IR videos while the other 5 classes are usually captured in constant positions. However, if we just rotate these 3 classes and keep the other 5 classes non-rotated, quantity of each class will be quite different, which is unacceptable for the results, thus all the test images are rotated in our experiment.

As is mentioned in section 1, there is another way to solve the rotation issue—data augmentation for training samples, so we set it as comparison. Besides, among all the  $l_1$ -norm minimization algorithms, OMP is the most widely used and homotopy is a typical fast convergence algorithm, so we test both of them. In a word, we compare our method (Algorithm 4) to three other HOG-SRC based methods—data augmentation + OMP (Algorithm 1), data augmentation + homotopy (Algorithm 2), main orientation rotation + OMP (Algorithm 3). In Alg. 1 and Alg. 2, the dictionary is a  $324 \times 14400$  matrix because data augmentation is processed for all the training images. In addition, we also compare KNN and SVM to our algorithm as KNN and SVM are the very typical small sample learning algorithms.

In this experiment, we firstly resize all the test images to  $40 \times 40$  pixels. Main orientations of test images are

computed afterwards. As is the same with dictionary learning procedure, we choose  $0^\circ$  to be the reference direction. Then each of the test images is rotated until its main orientation coincides with the reference direction. After that, we extract HOG features of these test images, which are also 324-dimension vectors. For each HOG feature, its corresponding sparse representation coefficients is computed according to Eq. (12). At last, identity of this target would be predicted according to Eq. (14). In sparse coefficients computing process, we set  $\gamma$  to  $10^{-5}$  and tolerance to 0.005. The stop criterion is either residual error is smaller than tolerance or the number of iterations reaches 1000. Besides, for KNN method, K is set to 3. For SVM method, linear function is chosen to be the kernel function. All the experiments are repeated 10 times and for each experiment, training and testing samples are selected randomly and independently. Identification rate is defined as Eq. (15). Sample test images and their sparse representation coefficients are shown in Fig. 7. The average results are shown in Table 3.

$$\text{identification rate} = \left( \sum_{\text{repeat}=1}^{10} \frac{\text{number of right identified targets}}{\text{number of targets to be identified}} \right) / 10 \quad (15)$$

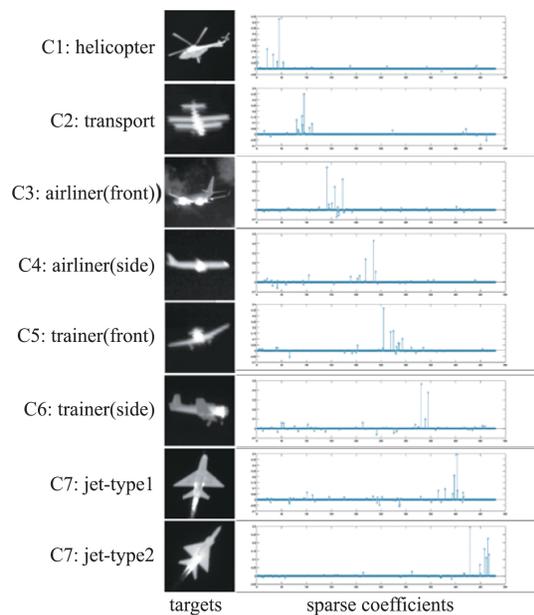


Fig. 7 Sample test images and their sparse representation coefficients.  
图7 部分测试图像及相应的稀疏表示系数

Figure 7 illustrates that sparse coefficients for different classes mainly concentrate on their respective regions. This phenomenon verifies the inherent quality of the sparsest representation—among all subsets of basic atoms, the sparsest solution selects the subset which most compactly expresses the input signal and rejects all other possible but less compact representations. Table 3 shows that methods based on SRC exhibit the best overall. In comparison of Alg. 2 and Alg. 4 (or Alg. 1 and Alg. 3), the size of dictionary has little impact on identi-

fication accuracy. Comparing OMP and homotopy convergence, we see that homotopy performs much better. Besides, identification frames per second of Alg. 4 is up to 82.6FPS, this is enough for aerial identification task. As to KNN and SVM, it can be seen that KNN also performs well that it predicts all types beyond 90% and even has the highest identification accuracy for trainer(side)-type. On the contrary, SVM performs the worst, this might be improved by using other kernel functions. Nevertheless, SVM is essentially a binary classifier so it solves multi-classification task not that efficiently.

#### 2.4 Experimental results for anti-noise capability

On the basis of fundamental physics, every object at any given absolute temperature above 0K emits thermal radiation including atmosphere<sup>[24]</sup>, thus one of the most particular characters of IR images is low SNR. To validate how the proposed method performs under noise influence, we randomly choose a number of pixels in each test image to be corrupted. The percentage is from 10 percent to 90 percent and the corruption is done by adding the original intensity with independent and identically distributed samples from a Gaussian distribution. Fig. 8 shows several sample test images and the experimental results are shown in Fig. 9.

We see that our algorithm recovers the identity of all targets beyond 80% accuracy even when the test images are corrupted at 50%. This performance is due to the in-

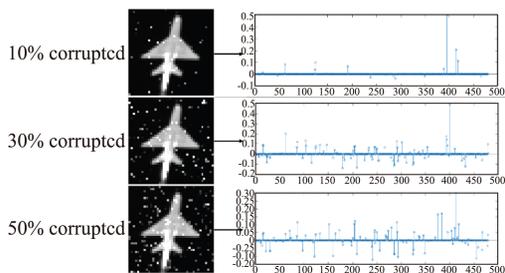


Fig. 8 Sample test images for anti-noise capability. Left row: sample test images with percent corrupted. Right row: sparse representation coefficients of the test images

图8 对测试图像叠加噪声后的稀疏表示系数. 左列: 叠加噪声后的测试图像. 右列: 左侧图像对应的稀疏表示系数

herent property of sparse representation—when the test image  $\mathbf{y}$  is partially corrupted, function (7) should be modified as

$$\mathbf{y} = \mathbf{y}_0 + \text{error} = \mathbf{D}\mathbf{x} + \text{error} \quad (16)$$

According to Wright, SRC is proved to remain 100% recognition rate even when the image is corrupted by 60%. In our experiment, main orientation HOG with SRC is unable to reach that high accuracy. This is owing to the feature extraction process in our method. When dictionary is constructed by pixels, identification may still be available on the basis of remaining pixels even if some pixels are corrupted. In contrast, HOG descriptor is based on gradient image, which is more sensitive to noises. As we can see in Fig. 9, although our method does not show that strong anti-noise capability in comparison with SRC in Ref[10], comparing to KNN or SVM methods, it still recovers the identity of all targets beyond 80% accuracy when the test images are corrupted at 50%.

Comparing OMP and homotopy convergence, it can be seen that homotopy is much more robust to noise. In comparison of Alg. 2 and Alg. 4 (or Alg. 1 and Alg. 3), it can be seen that the increasement of dictionary size spoils the identity, this is because in convergence process, proper stopping criterion should be set. When dictionary size is small,  $l_1$ -minimization problem may easily fall to converge to the global optimum. On the contrary, when dictionary size enlarges too much, convergence process may stop before converging to the optimum, and this will become worse when noises added in.

### 3 Conclusion

In this paper, we presented a fast rotation-invariant identification algorithm based on HOG descriptor and SRC classifier. The key idea to rotation invariance is that in IR images, aeroengine shows the strongest thermal radiation, so the main orientation of a target can be computed from gradient information. Experiment results demonstrate that our method appears not only high identification accuracy but also robustness to noise. In future, we

Table 3 Identification rates of various methods on the 8 classes of aerial targets.

表3 不同算法下的分类正确率

	Alg. 1	Alg. 2	Alg. 3	Alg. 4	Alg. 5	Alg. 6
C1: helicopter	0.956	0.984	0.953	<b>0.986</b>	0.955	0.833
C2: transport	0.915	<b>0.981</b>	0.903	0.973	0.936	0.971
C3: airliner(front)	0.974	0.982	0.974	<b>0.986</b>	0.968	0.867
C4: airliner(side)	0.972	<b>0.991</b>	0.965	<b>0.991</b>	0.907	0.846
C5: trainer(front)	0.891	<b>0.990</b>	0.898	0.985	0.948	0.961
C6: trainer(side)	0.908	0.978	0.919	0.981	<b>0.983</b>	0.971
C7: jet-type1	0.956	0.968	0.927	<b>0.988</b>	0.954	0.881
C8: jet-type2	0.795	0.976	0.832	<b>0.978</b>	0.961	0.763
average identification accuracy	0.921	0.981	0.921	0.983	0.951	0.887

Alg.1—data augmentation+HOG-SRC+OMP; Alg.2—data augmentation+HOG-SRC+homotopy; Alg.3—main orientation HOG-SRC+OMP; Alg.4—main orientation HOG-SRC+homotopy; Alg.5—main orientation HOG-KNN; Alg.6—main orientation HOG-SVM.

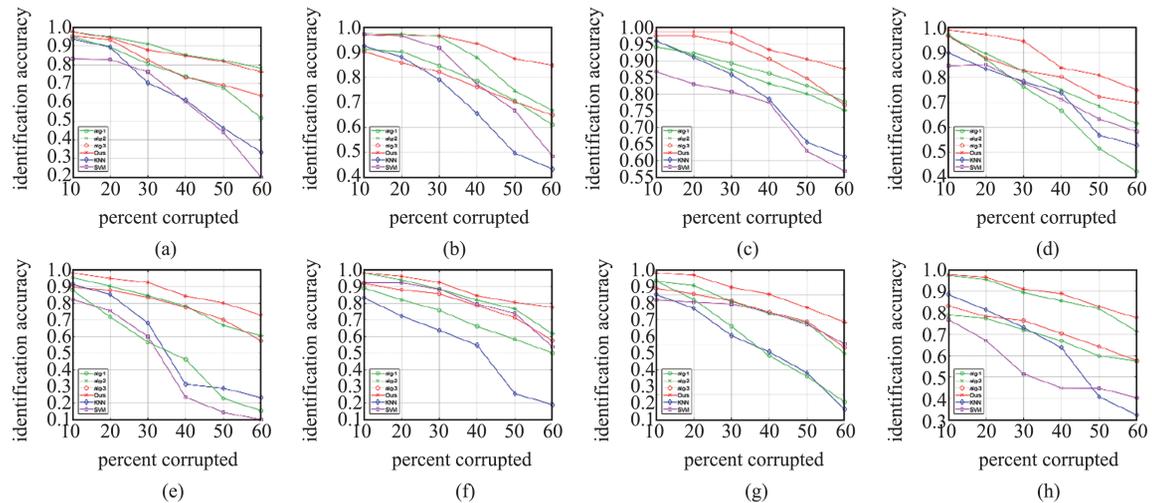


Fig.9 Identification accuracy when test images are percent corrupted. (a) helicopter-type identification accuracy; (b) transport-type identification accuracy; (c) airliner (front)-type identification accuracy; (d) airliner (side)-type identification accuracy; (e) trainer (front)-type identification accuracy; (f) trainer (side)-type identification accuracy; (g) jet-type1 identification accuracy; (h) jet-type2 identification accuracy

图9 噪声叠加对分类正确率的影响. (a)直升机; (b)运输机; (c)民航(正面); (d)民航(侧面); (e)教练机(正面); (f)教练机(侧面); (g)喷气式飞机(型号I); (h)喷气式飞机(型号II)

plan to further expand our dataset from two aspects: on one hand, more types of aerial targets are expected to be added. On the other hand, quantity of each type is to be enlarged, especially for those in cloudy sky, as thus validation of targets in complex background can be completed.

## Acknowledgment

This work was supported by the Thirteen Five National Defense Research Foundation (Jzx2016-0404/Y72-2), and Shanghai Key Laboratory of Criminal Scene Evidence funded Foundation (2017xcwzk08).

## Reference

- [1] Lowe D G. Distinctive image features from scale-invariant keypoints[J]. *International journal of computer vision*, 2004, **60**(2): 91-110.
- [2] Bay H, Ess A, Tuytelaars T, *et al.* Speeded-up robust features (SURF) [J]. *Computer vision and image understanding*, 2008, **110**(3): 346-359.
- [3] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C]. international Conference on computer vision & Pattern Recognition (CVPR'05). IEEE Computer Society, 2005, 1: 886--893.
- [4] Fei-Fei L, Perona P. A bayesian hierarchical model for learning natural scene categories [C]. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). IEEE, 2005, **2**: 524-531.
- [5] Huang G B, Lee H, Learned-Miller E. Learning hierarchical representations for face verification with convolutional deep belief networks [C]. 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012: 2518-2525.
- [6] Takacs G, Chandrasekhar V, Tsai S, *et al.* Unified real-time tracking and recognition with rotation-invariant fast features [C]. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2010: 934-941.
- [7] Chen J, Takiguchi T, Ariki Y. Rotation-reversal invariant HOG cascade for facial expression recognition [J]. *Signal, Image and Video Processing*, 2017, **11**(8): 1485-1492.
- [8] Liu B, Wu H, Su W, *et al.* Rotation-invariant object detection using Sector-ring HOG and boosted random ferns [J]. *The Visual Computer*, 2018, **34**(5): 707-719.
- [9] Liu B, Wu H, Su W, *et al.* Sector-ring HOG for rotation-invariant human detection [J]. *Signal Processing: Image Communication*, 2017, **54**: 1-10.
- [10] Wright J, Yang A Y, Ganesh A, *et al.* Robust face recognition via sparse representation [J]. *IEEE transactions on pattern analysis and machine intelligence*, 2009, **31** (2): 210-227.
- [11] Peng Y, Li L, Liu S, *et al.* Space - frequency domain based joint dictionary learning and collaborative representation for face recognition [J]. *Signal Processing*, 2018, **147**: 101-109.
- [12] Zeng S, Gou J, Yang X. Improving sparsity of coefficients for robust sparse and collaborative representation-based image classification [J]. *Neural Computing and Applications*, 2018, **30**(10): 2965-2978.
- [13] Yang S, Wen Y. A novel SRC based method for face recognition with low quality images [C]. 2017 IEEE International Conference on Image Processing (ICIP). IEEE, 2017: 3805-3809.
- [14] Engan K, Aase S O, Husøy J H. Multi-frame compression: Theory and design [J]. *Signal Processing*, 2000, **80** (10): 2121-2140.
- [15] Cai S, Weng S, Luo B, *et al.* A dictionary-learning algorithm based on method of optimal directions and approximate K-SVD [C]. 2016 35th Chinese Control Conference (CCC). IEEE, 2016.
- [16] Aharon M, Elad M, Bruckstein A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation [J]. *IEEE Transactions on signal processing*, 2006, **54**(11): 4311.
- [17] Lu Z, Zhang L. Face recognition algorithm based on discriminative dictionary learning and sparse representation [J]. *Neurocomputing*, 2016, **174**: 749-755.
- [18] Yang M, Zhang L, Feng X, *et al.* Fisher discrimination dictionary learning for sparse representation [C]. 2011 International Conference on Computer Vision. IEEE, 2011:

- 543–550.
- [19] Figueiredo M A T, Nowak R D, Wright S J. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems[J]. *IEEE Journal of selected topics in signal processing*, 2007, **1**(4): 586–597.
- [20] Turkyilmazoglu M. An effective approach for evaluation of the optimal convergence control parameter in the homotopy analysis method[J]. *Filomat*, 2016, **30**(6): 1633–1650.
- [21] Wright S J, Nowak R D, Figueiredo M A T. Sparse reconstruction by separable approximation[J]. *IEEE Transactions on Signal Processing*, 2009, **57**(7): 2479–2493.
- [22] Beck A, Teboulle M. A fast iterative shrinkage–thresholding algorithm for linear inverse problems[J]. *SIAM journal on imaging sciences*, 2009, **2**(1): 183–202.
- [23] Yang J, Zhang Y. Alternating direction algorithms for  $l_1$ -problems in compressive sensing[J]. *SIAM journal on scientific computing*, 2011, **33**(1): 250–278.
- [24] Vollmer M, Möllmann K P. Infrared thermal imaging: fundamentals, research and applications[M]. John Wiley & Sons, 2017.