

# Infrared Target Recognition using Heterogeneous Features with Multi-kernel Transfer Learning

**Xin Wang<sup>1\*</sup>, Xin Zhang<sup>1</sup> and Chen Ning<sup>2</sup>**

<sup>1</sup>College of Computer and Information, Hohai University  
Nanjing, Jiangsu 211100 - China  
[e-mail: wang\_xin@hhu.edu.cn]

<sup>2</sup>School of Physics and Technology, Nanjing Normal University  
Nanjing, Jiangsu 210023 - China

\*Corresponding author: Xin Wang

*Received November 14, 2019; revised February 24, 2020; accepted August 3, 2020;  
published September 30, 2020*

---

## Abstract

Infrared pedestrian target recognition is a vital problem of significant interest in computer vision. In this work, a novel infrared pedestrian target recognition method that uses heterogeneous features with multi-kernel transfer learning is proposed. Firstly, to exploit the characteristics of infrared pedestrian targets fully, a novel multi-scale monogenic filtering-based completed local binary pattern descriptor, referred to as MSMF-CLBP, is designed to extract the texture information, and then an improved histogram of oriented gradient-fisher vector descriptor, referred to as HOG-FV, is proposed to extract the shape information. Second, to enrich the semantic content of feature expression, these two heterogeneous features are integrated to get more complete representation for infrared pedestrian targets. Third, to overcome the defects, such as poor generalization, scarcity of tagged infrared samples, distributional and semantic deviations between the training and testing samples, of the state-of-the-art classifiers, an effective multi-kernel transfer learning classifier called MK-TrAdaBoost is designed. Experimental results show that the proposed method outperforms many state-of-the-art recognition approaches for infrared pedestrian targets.

---

**Keywords:** Infrared, target recognition, heterogeneous features extraction, multi-kernel learning, transfer learning

---

This work was supported in part by the Fundamental Research Funds for the Central Universities (Grant No. 2019B15314), in part by Jiangsu Province Government Scholarship for Studying Abroad, in part by the Six Talents Peak Project of Jiangsu Province (Grant No. XYDXX-007), and in part by the 333 High-Level Talent Training Program of Jiangsu Province.

## 1. Introduction

**P**edestrian target recognition in infrared (IR) images is an important research branch of infrared image processing. It is used in various applications such as video surveillance, intelligent transportation, and human-computer interaction. However, robust pedestrian target recognition in infrared images is actually not an easy task. Since the pedestrian targets are usually drawn in background clutters so that they have inapparent appearances. Also, the pedestrian targets are not rigid targets. Their postures may change over time. Therefore, effective and robust recognition of infrared pedestrian targets is a challenging task [1].

Generally, infrared target recognition consists of two key modules: target feature extraction and classification. A number of literatures have designed various feature extraction approaches for infrared objects. For instance, in [2], a histogram of oriented gradient (HOG) was used to dig out the shape information. In [3], a gradient location orientation histogram (GLOH) was presented to extract the gradient features. In [4], a modified local binary pattern (LBP) was proposed to extract the texture feature for infrared targets. In [5], an intensity self-similarity (ISS) measure was adopted for intensity feature extraction. As illustrated above, these literatures merely extract a certain kind of features for infrared targets, which is unable to describe target characteristics completely. Therefore, some literatures have proposed to extract different features and fuse them together to improve the recognition performance. For example, in [6], HOG and LBP descriptors were used to extract shape and texture features respectively and then integrated together for night time pedestrian classification. In [7], for efficient pedestrian detection, the gradient and phase congruency concepts were proposed to capture the shape features, and a center-symmetric local binary pattern approach was used to capture the texture of the image. In [8], four features, including local binary pattern, Gabor jet descriptor, Weber local descriptor and down-sampling feature were combined for thermal target recognition. Through the results of methods that use combined features for target recognition, we find that, first, extracting various features and fusing them together can boost the capability of infrared target recognition compared with using a single kind of feature. Second, although some literatures have proposed to extract different kinds of features for infrared object recognition, these features are sometimes affiliated to the same kind so that the description of the target is still not complete. What's more, most features designed for pedestrian targets are low-level features, lacking of important semantic information, which limits the further promotion of recognition performance.

As for the classifiers for target recognition, the widely used ones include support vector machine (SVM) [9], Adaboost [10], K-nearest neighbor (KNN) [11], sparse representation (SR) [12], etc. These classifiers can only work well under two conditions: (1) the training and testing samples are drawn from the same feature space and distribution; (2) there exist sufficient training samples to train an effective classifier. Nevertheless, it is not easy to satisfy both of these two conditions in practical applications. First, due to the special imaging mechanism of infrared images, infrared targets may suffer large differences, resulting in huge deviations in the distributions and semantic contents between training and testing samples. Second, the available labeled infrared samples are usually scarce, and collecting a large number of new and effective labeled samples will take a big price. Thus, it is of great significance to make full use of a small number of labeled training samples to establish a reliable classifier for recognition.

In recent years, transfer learning [13] has attracted increasing attention in the field of machine learning. It relaxes the above two conditions and can well handle the classifier training problem with only a small amount of labeled data in the target domain by transferring the existing useful knowledge. At present, transfer learning has been successfully applied to many machine learning fields, such as natural language processing, text classification and target detection [14-16]. For instance, in [14], an iterative reweighting heterogeneous transfer learning framework was designed for remote sensing image classification. In [15], a target recognition method for synthetic aperture radar was built via transfer learning from simulated data. In [16], a transferrable representation learning model was proposed to enhance face recognition performance. For infrared pedestrian target recognition problem, to make full use of transfer learning, we construct an improved multi-kernel transfer learning classifier with the fusion of heterogeneous features.

The proposed method has two main contributions as follows: (1) To overcome the defects of incomplete information extracted by using one single feature extraction method, two enhanced feature extraction schemes are proposed to extract two kinds of heterogeneous features, i.e., MSMF-CLBP and HOG-FV. Thus, more complete description about the targets can be gotten. Moreover, different from the low-level features, these two kinds of features belong to higher level features that can effectively reflect the semantic information. (2) For target recognition, an effective multi-kernel transfer learning classifier called MK-TrAdaBoost is designed. Compared with the traditional transfer learning classifier, the proposed classifier can effectively solve the problem of lack of sufficient labeled infrared samples and at the same time, enhance distinguishability of the data to be classified. Hence, it can achieve much better recognition results.

The rest of this paper is organized as follows. In Section 2, the proposed method is introduced in detail. Experimental results are presented in Section 3. Conclusions are finally drawn in Section 4.

## 2. Proposed Method

The proposed infrared pedestrian target recognition method is presented in this section. The overall framework is depicted in Fig. 1. It consists of two stages: the training stage and the testing stage. In the training stage, two different kinds of training sets, i.e., source training set and auxiliary training set are firstly constructed with infrared sample images as well as visible sample images (both pedestrian and non-pedestrian targets included). Then, two heterogeneous features are extracted from all sample images. One feature is a novel multi-scale monogenic filtering-based completed local binary pattern (referred to as MSMF-CLBP), and the other is an improved histogram of oriented gradient-fisher vector descriptor (referred to as HOG-FV). Based on these two enhanced features, an effective multi-kernel transfer learning classifier called MK-TrAdaBoost is designed. In the testing stage, given an infrared image to be recognized, the two heterogeneous features are firstly extracted. Then, by using the trained classifier, MK-TrAdaBoost, the target can be well recognized.

### 2.1 Heterogeneous Features Extraction

The heterogeneous features to be proposed are MSMF-CLBP and HOG-FV that respectively describe the texture and shape information of infrared targets. Their combination can not only reduce the effective information loss in the image, but also enrich the semantic content of feature expression, which is beneficial for the subsequent recognition.

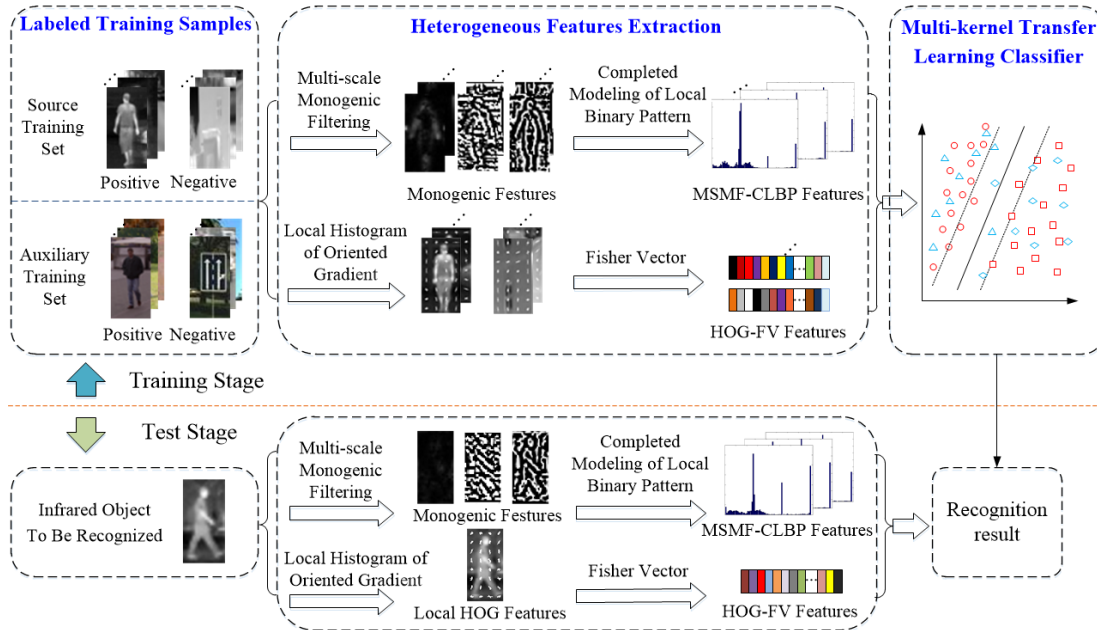


Fig. 1. Framework of the proposed method.

### 2.1.1 MSMF-CLBP Feature Extraction

As is known, the completed modeling of LBP feature (CLBP) is the development of the classic LBP feature. Compared with LBP, CLBP has been proved to be more efficient [17]. In [18], to further enhance the descriptive ability of CLBP, a multi-scale CLBP feature, called MS-CLBP, was proposed to characterize the texture information in the image for land-use scene classification. Although CLBP and its variant MS-CLBP are useful to describe the image texture characteristics, for infrared targets, they are very sensitive to noise and illumination changes, since they are generally calculated based on the image intensities. Therefore, to overcome this problem, in this work, we propose a novel multi-scale monogenic filtering-based CLBP descriptor, named MSMF-CLBP, to extract the texture features.

In recent years, a Riesz transform based generalization of two-dimensional (2-D) analytic signal was addressed in [19]. Based on it, a more sophisticated analytic signal called the monogenic signal was produced, which owns two crucial advantages. First of all, it can represent signals compactly almost without information loss. Second, it orthogonally decomposes the original signal into three components: local phase, local amplitude and local orientation, providing effectual solutions to many 2-D signal processing problems [20, 21]. Based on these merits, here we proposed to employ the monogenic signal filtering on the target image to be recognized before implementing the CLBP feature extraction to reduce the influences of noise and other negative factors.

Given an infrared pedestrian target image  $I(l)$ , where  $l = (x, y)^T$  denotes the spatial domain coordinate, a multi-scale monogenic filtering (MSMF) scheme is first applied to it. Suppose  $I_R(l)$  is the Riesz-transformed image of  $I(l)$ . Combining  $I(l)$  and  $I_R(l)$  as the following form:

$$I_M(l) = I(l) - (i, j) I_R(l) \quad (1)$$

then we can get the monogenic signal  $I_M(l)$ . In (1),  $i$ ,  $j$  represent the imagery units. Based on  $I_M(l)$ , three components, local amplitude, phase, and orientation, can be calculated by:

$$A(l) = \sqrt{I(l)^2 + |I_R(l)|^2} \quad (2)$$

$$\varphi(l) = \text{atan}(|I_R(l)|, I(l)) \in (-\pi, \pi] \quad (3)$$

$$\theta(l) = \text{atan}(I_y(l)/I_x(l)) \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right] \quad (4)$$

where  $A(l)$  is the local amplitude, reflecting the local energetic information.  $\varphi(l)$  and  $\theta(l)$  are the local phase and orientation, reflecting the local structural and geometric information, respectively.  $I_x(l)$  is the i-imaginary component of  $I_M(l)$ , while  $I_y(l)$  is the j-imaginary component of  $I_M(l)$ .

Since the infrared pedestrian image is of finite length, it results in infinite spectra in the frequency domain. To capture broad spectral information with compact support, we adopt the Log-Gabor filter to extend the image to be infinite. Then, the monogenic signal in (1) is rewritten as:

$$I_M(l) = (I(l) * h_{LG}(l)) - (i, j)(I_R(l) * h_{LG}(l)) \quad (5)$$

where  $h_{LG}$  is the Log-Gabor kernel [22].

Then, to enhance the robustness of the monogenic signal against noise, directional changes, etc, we propose to tune the scale of the Log-Gabor filter to generate the monogenic signal at different scales  $I_M^k$  ( $k = 1, \dots, S$ ), where  $S$  is the total number of scales. Consequently, we can get the components of monogenic signal at different scale-space  $A^k$ ,  $\phi^k$ , and  $\theta^k$  ( $k = 1, \dots, S$ ), representing local amplitude, phase, and orientation of  $I_M^k$ , respectively.

Subsequently, the CLBP operator is performed on the multi-scale monogenic filtering results  $A^k$ ,  $\phi^k$ , and  $\theta^k$  ( $k = 1, \dots, S$ ) to obtain the multi-scale monogenic filtering-based completed local binary pattern features, which can effectively and robustly describe the texture information. Specifically, for each of monogenic filtering result, three operators, namely CLBP\_S, CLBP\_M and CLBP\_C [17], are utilized to code its C, S, and M features, respectively. Given a pixel in the result image, the codes are computed by comparing it with its neighbors:

$$CLBP\_S_{B,E} = \sum_{b=0}^{B-1} s(g_b - g) 2^b, \quad s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (6)$$

$$CLBP\_M_{B,E} = \sum_{b=0}^{B-1} t(m_b, \varsigma) 2^b, \quad t(x, \varsigma) = \begin{cases} 1, & x \geq \varsigma \\ 0, & x < \varsigma \end{cases} \quad (7)$$

$$CLBP\_C_{B,E} = t(g_b, \varsigma_I) \quad (8)$$

where  $g$  is the value of the central pixel.  $g_b$  is the value of its neighbors.  $B$  is the total number of involved neighbors.  $E$  is the radius of the neighborhood.  $m_b = |g_b - g|$ .  $\varsigma$  is a

threshold to be determined adaptively.  $\zeta_I$  is a threshold that is usually set as the average value of the whole image. CLBP\_S, CLBP\_M and CLBP\_C construct the completed LBP framework. If they are combined together, more discriminative features can be gotten to represent the image. In this paper, in order to ensure higher accuracy of feature extraction and lower feature dimension, CLBP\_M and CLBP\_C are firstly combined to build a 2-D joint histogram CLBP\_M/C. Then, the histogram is converted to a 1-D histogram, which is subsequently concatenated with CLBP\_S to generate a joint histogram, denoted by CLBP\_S\_M/C.

At last, by applying the multi-scale monogenic filtering and completed modeling of local binary pattern operator to the original image successively, the integration results ensure that, the novel feature (referred to as MSMF-CLBP) is more robust to interferences compared to the traditional CLBP as well as MS-CLBP.

### 2.1.2 HOG-FV Feature Extraction

In view of histogram of oriented gradient's robustness to noise and changes in local shape, HOG has been widely used in target recognition, tracking, etc. It is efficient to utilize HOG features to describe the shape information of images, but they essentially belong to low-level features. To enhance their ability to describe images, feature coding becomes a feasible means. Bag-of-words (BoWs) is a famous model to encode HOG features for enriching their semantic content [23]. The BoWs model builds a dictionary by selecting a large number of visual vocabularies, and then the dictionary is utilized to encode the low-level features. However, the performance of BoWs is extremely dependent on the size of the dictionary and it generally requires a dictionary with large size to ensure the good performance. In contrast, fisher vector (FV), as an extension of BoWs, has looser demands for dictionaries. It can achieve good results based on a dictionary with smaller size, which is at the same time beneficial to reducing the time complexity [24]. Hence, in this work, we propose to encode the low-level HOG features through FV instead of BoWs to get more powerful shape features. The specific process is described as follows.

Given an image  $I$ , the local HOG features are extracted by using gradient histogram [25]. At first, the lateral gradient  $H(l)$  and longitudinal gradient  $V(l)$  of each pixel  $l$  are calculated with the gradient operator  $[-1, 0, 1]$ , respectively. Then, the gradient magnitude  $M(l)$  and orientation  $O(l)$  of the pixel can be gotten by:

$$M(l) = \sqrt{H(l)^2 + V(l)^2} \quad (9)$$

$$O(l) = \text{atan}[V(l)/H(l)] \quad (10)$$

Subsequently, divide  $I$  into a number of cells, each of which is divided into 9 gradient orientations. As a result, we can get a 9-dimensional vector. By putting 4 adjacent cells into one block, a 36-dimensional HOG block vector can be obtained. After that,  $\ell_2$  norm is applied to the block vector and the principal component analysis (PCA) is used to reduce its dimension to 30 so that the dimension of feature after fisher vector encoding is not too high. Compared with the classic HOG feature extraction method, this local HOG feature extraction method does not adopt sliding fusion on the block vector, so it can fully preserve the local gradient features of the image.

Next, fisher vector coding, which is implemented with the Gaussian Mixture Model (GMM), is applied to the extracted local HOG features. Suppose there exist  $T$  local HOG features to

be encoded for the image, i.e.,  $X = \{x_t, t=1,2,\dots,T\}$ , where the dimension of  $x_t$  is  $D$ .  $\lambda = \{w_i, \mu_i, \sigma_i, i=1,2,\dots,N\}$  denotes the parameter set of GMM, where  $w_i$ ,  $\mu_i$ ,  $\sigma_i$  represent the weight, mean and covariance of  $i$ -th Gaussian kernel in FV. Each Gaussian kernel stands for a visual vocabulary of the dictionary. Suppose  $x_t (t=1,2,\dots,T)$  satisfies the independent and identical distribution, then

$$L(X|\lambda) = \log p(X|\lambda) = \sum_{t=1}^T \log p(x_t|\lambda) \quad (11)$$

where the likelihood that  $x_t (t=1,2,\dots,T)$  can be generated by GMM is:

$$p(x_t|\lambda) = \sum_{i=1}^N w_i p_i(x_t|\lambda), \quad \sum_{i=1}^N w_i = 1 \quad (12)$$

And the occupancy probability that  $x_t$  is generated by the  $i$ -th Gaussian kernel is computed by:

$$\gamma_t(i) = \frac{w_i p_i(x_t|\lambda)}{\sum_{j=1}^N w_j p_j(x_t|\lambda)} \quad (13)$$

Since less information is carried by the gradient vector obtained from  $w_i$ , the effect of  $w_i$  is usually ignored. Then, by taking the partial derivative of  $L(X|\lambda)$ , the gradient vectors for  $\mu_i$  and  $\sigma_i$  can be provided as follows:

$$G_{\mu,i}^X = \frac{1}{T\sqrt{w_i}} \sum_{t=1}^T \gamma_t(i) \left( \frac{x_t - \mu_i}{\sigma_i} \right) \quad (14)$$

$$G_{\sigma,i}^X = \frac{1}{T\sqrt{2w_i}} \sum_{t=1}^T \gamma_t(i) \left[ \frac{(x_t - \mu_i)^2}{\sigma_i^2} - 1 \right] \quad (15)$$

Finally, the encoded local HOG features can be described as  $G_\lambda^X = [G_{\mu,i}^X, G_{\sigma,i}^X]$  with the dimension of  $2 \times N \times D$ . Since FV coding utilizes higher order statistics (mean and covariance) to reduce quantization error so that the encoded features have lower information loss while depicting the image information. At last, by connecting local HOG features with fisher vector coding, the low-level local HOG features of the image are transformed into higher level features with rich semantic information. Thus, the ability of features to describe the image is enhanced, which has positive effect on the subsequent recognition performance. The improved histogram of oriented gradient-fisher vector descriptor is referred to as HOG-FV in the work.

## 2.2 Improved Multi-kernel Transfer Learning Classifier

### 2.2.1 TrAdaBoost

TrAdaBoost is essentially an instance-based transfer learning algorithm. Suppose there exist a limited set of labeled training samples, called source training set, which are subject to the same distribution as the test samples. Nevertheless, due to its limited number, it is impossible to train a good classifier. Fortunately, there is another set of samples, called auxiliary training set,



which has a different distribution with the test samples. The idea of TrAdaBoost is to adopt a boosting technique [26] to seek appropriate samples from the auxiliary training set, and transfer them to the training and learning process of the source training data.

Suppose the source training set is  $X_s$ , and the auxiliary training set is  $X_d$ . The label is  $L = \{+1, -1\}$ . The whole training set can be represented by:  $X = X_s \cup X_d$ . The labeled source and auxiliary training sets are respectively denoted as:

$$T_s = \left\{ \left( x_j^s, l_j^s \right), x_j^s \in X_s, l_j^s \in L, j = 1, 2, \dots, m \right\} \quad (16)$$

$$T_d = \left\{ \left( x_k^d, l_k^d \right), x_k^d \in X_d, l_k^d \in L, k = 1, 2, \dots, n \right\} \quad (17)$$

where  $m$  and  $n$  are the number of source and auxiliary training samples, respectively. By using appropriate weight-adjustment strategies, TrAdaBoost can train an excellent classifier based on the limited samples in  $T_s$  and the valid samples in  $T_d$ .

## 2.2.2 Multi-kernel Learning

As is known, the kernel function method is effective to solve the problem of pattern analysis. It maps the input data into a new feature space through nonlinear mapping so as to get a more discriminative feature representation. However, in the case of heterogeneous samples or uneven sample distributions, it is unreasonable to apply a single kernel to all the samples. Therefore, an idea of multi-kernel learning (MKL) is derived, which integrates multiple basic kernel functions into a unified framework [27]. Then, in the new feature space, the data can be better expressed and the distinguishability of features can be well enhanced. The multi-kernel learning model is described as:

$$K = \sum_{i=1}^k \alpha_i \kappa_i, \quad \alpha_i \geq 0, \quad \sum_{i=1}^k \alpha_i = 1 \quad (18)$$

where  $k$  is the number of basic kernel functions.  $\kappa_i$  and  $\alpha_i$  reveal respectively the  $i$ -th basic kernel function and its weight coefficient. A popular kernel, called Gaussian radial basic function (RBF), is employed as the basic kernels in this paper, for it can handle the non-linear mapping between the class labels and features robustly [28]. A RBF is defined as:

$$\kappa(x_i, x_j) = \exp \left( -\frac{\|x_i - x_j\|^2}{2\sigma^2} \right) \quad (19)$$

where  $\sigma$  is the radial width, indicating different scales of kernel functions.

A number of RBF kernels are then integrated to construct a multi-scale kernel by using the following procedures. Firstly, initialize the range of  $\sigma : [\sigma_{\min}, \sigma_{\max}]$ , and select  $k$  RBF kernel functions with different scales:  $\sigma_{\min} \leq \sigma_1 < \sigma_2 < \dots < \sigma_k \leq \sigma_{\max}$ . Second, calculate  $\sigma_m = \sqrt{d/2}$ , where  $d$  denotes the dimension of features to be classified. Third, calculate the distances between  $\sigma_i (i = 1, 2, \dots, k)$  and  $\sigma_m$ , and adjust the corresponding kernel function weight coefficient coefficients  $\alpha_i (i = 1, 2, \dots, k)$  based on the distances. Note that the smaller the distance is, the larger the corresponding coefficient is. Finally, the multi-scale kernel  $K$  can be obtained from the determined kernel functions with the corresponding coefficients.



### 2.2.3 Multi-kernel TrAdaBoost Classifier

TrAdaBoost achieves the transfer of valid samples by adjusting the weights of training samples in the iterative process, and ultimately combines the weak classifiers in each iteration into a strong classifier. In this paper, to improve the performance of the traditional TrAdaBoost, the idea of multi-kernel learning is embedded into the TrAdaBoost framework. First, support vector machine (SVM) is selected as the initial weak classifier. Then, the multi-scale kernel  $K$  is integrated into the SVM framework. Based on each iteration in TrAdaBoost, a more effective classifier can be learned. The improved algorithm in this paper is referred to as: Multi-kernel TrAdaBoost classifier (MK-TrAdaBoost). The detailed process of MK-TrAdaBoost is shown in [Table 1](#).

**Table 1.** MK-TrAdaBoost algorithm.

MK-TrAdaBoost algorithm	
Input: Source training set and auxiliary training set: $T_s$ and $T_d$ . The number of iterations: $N_I$	
Output: MK-TrAdaBoost classifier	
1. Initialize the related parameters of MK-TrAdaBoost, including the weight vectors:	
$\omega_i^1 = \begin{cases} \frac{1}{m}, & i = 1, 2, \dots, m \\ \frac{1}{n}, & i = m+1, m+2, \dots, m+n \end{cases}$ and the parameter $\beta = \frac{1}{1 + \sqrt{2 \ln \frac{m}{N_I}}}$ , where $m$ and $n$ are the number of source and auxiliary training samples, respectively.	
<b>for</b> $t = 1, 2, \dots, N_I$	
2. Normalize the weight vectors: $\omega_i^t = \frac{\omega_i^t}{\sum_{j=1}^{m+n} \omega_j^t}$ .	
3. Choose SVM as the initial classifier, and integrate $k$ RBF kernel functions with different radial width $\sigma$ to learn a multi-scale kernel as the kernel function of SVM. For the training set $X$ , obtain the corresponding class label mapping: $h_t : X \rightarrow Y$ , where $Y$ is the class label set.	
4. Calculate the recognition error rate of $h_t$ in $T_s$ : $\varepsilon_t = \sum_{i=1}^m \frac{\omega_i^t \cdot  h_t(x_i) - l_i^t(x_i) }{\sum_{i=1}^m \omega_i^t}$ , where $l_i^t(x_i)$ stands for the real class label of the training sample $x_i$ . Set the following rules: if $\varepsilon_t \geq 0.5$ , $\varepsilon_t$ is set to 0.499; if $\varepsilon_t = 0$ , $\varepsilon_t$ is set to 0.001.	
5. Compute $\beta_t = \frac{\varepsilon_t}{1 - \varepsilon_t}$ , and update the weights of all training samples by:	
$\omega_i^{t+1} = \begin{cases} \omega_i^t \cdot \beta_t^{ h_t(x_i) - l_i^t }, & i = 1, 2, \dots, m \\ \omega_i^t \cdot \beta_t^{ h_t(x_i) - l_i^t }, & i = m+1, m+2, \dots, m+n \end{cases}$	
<b>end</b>	
<b>return</b> $h_{N_I}$ .	

### 3. Experimental Results and Analysis

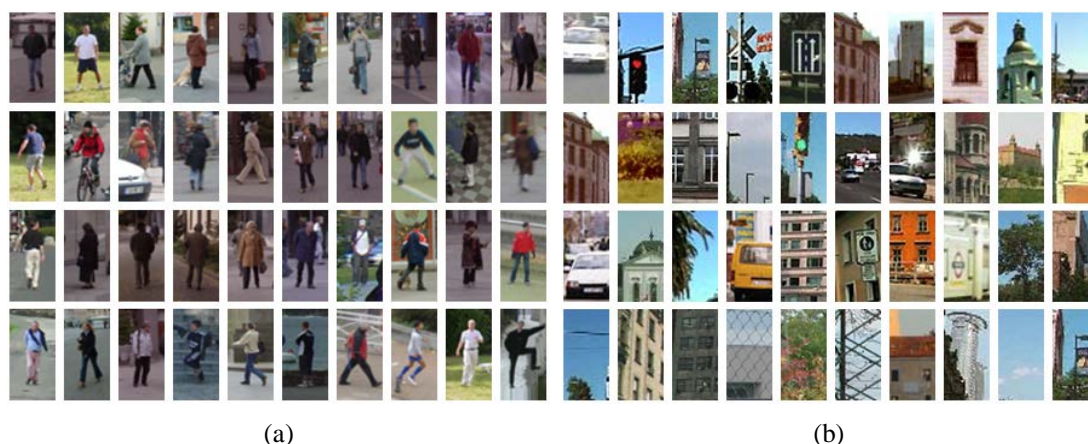
#### 3.1 Experimental Setup

To verify the proposed method, we do experiments on two famous datasets: LSI Far Infrared Pedestrian Dataset [29] and INRIA Person Dataset [30]. Thereinto, all infrared images in source training set and test set come from LSI Far Infrared Pedestrian Dataset that contains a lot of pedestrian images (positive) with various shapes and non-pedestrian images (negative) under different backgrounds and can meet the collection demands of training samples and test samples. We randomly select 310 positive images and 310 negative ones as source training samples from the LSI Far Infrared Pedestrian Dataset, and also randomly choose 500 positive images and 500 negative ones as test samples. The size of each sample image is  $64 \times 32$  pixels. Fig. 2 shows some examples of source training samples.



**Fig. 2.** Examples of source training samples. (a) Positive samples; (b) Negative samples.

The auxiliary training set is derived from INRIA Person Dataset, which contains 2416 visible pedestrian images with different postures and 1218 visible non-pedestrian images, such as streets, buildings and natural sceneries. It well satisfies the diversity required for the auxiliary samples in the experiments. From INRIA Person Dataset, 720 positive sample images and 720 negative ones are selected as the the auxiliary training sample set. Fig. 3 illustrates some examples of auxiliary training samples.



**Fig. 3.** Examples of auxiliary training samples. (a) Positive samples; (b) Negative samples.

All experiments are performed on a PC with Intel Core 1.5 GHz processor and 4.00 GB RAM. The simulation software is MATLAB R2014a. Three quantitative evaluations including accuracy rate (AR), F1-measure (F1), and standard deviation (SD) [31], are utilized to test the recognition performance. AR, standing for the proportion of correctly identified samples to the total number, is defined as:

$$AR = \frac{TP + TN}{TP + TN + FP + FN} \quad (20)$$

where  $TP$  and  $FN$  respectively denote the number of positive samples that are correctly identified and misidentified.  $TN$  and  $FP$  respectively represent the number of negative samples that are correctly identified and misidentified. F1 is defined as:

$$F_1 = \frac{2 \text{precision} * \text{recall}}{\text{precision} + \text{recall}} = \frac{2TP}{2TP + FN + FP} \quad (21)$$

where  $\text{precision} = TP / (TP + FP)$  is the recognition precision rate, while  $\text{recall} = TP / (TP + FN)$  is the recall rate. And F1 is a harmonic value for the precision and recall rates. In all, the higher the values of AR and F1 are, the better the recognition performance is. To strictly verify the proposed method, the experiments are carried out by multiple random tests, and the overall effectiveness is reflected by the average values of the above three indicators, i.e.,  $\overline{AR}$  and  $\overline{F_1}$ .

In addition, the stability of the presented method is verified by the standard deviation SD of the accuracy rate AR of multiple random tests, which is defined as:

$$SD = \sqrt{\frac{1}{N_t - 1} \sum_{i=1}^{N_t} \left( AR_i - \frac{1}{N_t} \sum_{i=1}^{N_t} AR_i \right)^2} \quad (22)$$

where  $N_t$  is the number of multiple random tests.  $AR_i$  is the accuracy rate of the  $i$ -th ( $i = 1, 2, \dots, N_t$ ) test. The smaller the value of SD is, the more stable the recognition performance is. It is worth pointing out that when doing each random test, the source training set and the test set are mixed at first, and then 400 positive images and 400 negative ones are selected separately as the source training samples and test samples. The auxiliary training samples keep unchanged.

Besides the above three indicators, the Receiver Operator Characteristic (ROC) curve as well as the area under the ROC curve (AUC) [32] are also broadly used. Therefore, we will summarize the experimental results with ROC curves and AUC values at last.

### 3.2 Evaluation of Heterogeneous Features Extraction

In this paper, our proposed method contains two main modules: heterogeneous features extraction, as well as the improved multi-kernel transfer learning classifier. Therefore, in this section, we first evaluate the former module by comparing it (referred to as MSMF-CLBP+HOG-FV) with several state-of-the-art feature extraction methods. First, to test the effectiveness of the proposed MSMF-CLBP feature, it is compared with the classic CLBP feature [17] and its variant MS-CLBP feature [18] that uses the Gabor filtering to reduce the noise interferences. Second, to test the effectiveness of the proposed HOG-FV feature, it is compared with the classic HOG feature [2] and its variant HOG-BOW feature [23] that uses the bag-of-words model to encode the HOG features. Third, the MSMF-CLBP feature and HOG-FV feature are respectively compared with the fused feature (i.e., MSMF-CLBP+HOG-FV) to demonstrate the effectiveness the heterogeneous features

integration. Fourth, our proposed MSMF-CLBP+HOG-FV is also compared two state-of-the-art features, the ISS feature [5] and the HOG+CLBP feature [33] which directly fuses the HOG feature with the CLBP feature. The comparison results are given in Table 2, Table 3 and Table 4. Thereinto, Table 2 illustrates the accuracy rate comparison with different features of 10 random tests. Table 3 shows the F1-measure comparison with different features of 10 random tests. The average performance comparison with different features is given in Table 4.

As can be seen from these tables, our proposed MSMF-CLBP feature is much more effective than CLBP and MS-CLBP, for the values of the two indicators, accuracy rate and F1-measure have increased by at least 1%, while the SD has decreased by at least 2%. Second, our proposed HOG-FV feature is consistently better than the reference feature, HOG and HOG-BOW. Third, the overall recognition performance is further boosted when using the fused MSMF-CLBP+HOG-FV feature compared with MSMF-CLBP or HOG-FV. The values of accuracy rate and F1-measure have increased by approximately 2%, while the SD has decreased by 2%. Fourth, the performance of our proposed feature is also better than those of ISS and HOG+CLBP. Although HOG+CLBP also contains two heterogeneous features, it lacks of high-level semantic content.

**Table 2.** Accuracy rate (%) comparison with different features.

Different features	Accuracy rates of 10 random tests									
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
MSMF-CLBP	98.63	<b>98.88</b>	98.25	98.75	98.75	98.75	98.75	98.13	98.63	98.63
CLBP	94.63	94.75	96.00	95.50	95.63	95.75	96.00	96.00	<b>96.13</b>	95.25
MS-CLBP	97.75	97.88	97.00	97.00	97.88	<b>98.25</b>	96.88	97.75	97.50	96.50
HOG-FV	<b>97.25</b>	96.38	96.25	96.75	96.88	95.88	97.25	95.88	96.38	96.75
HOG	95.00	95.75	95.38	96.00	96.25	96.13	96.00	<b>96.75</b>	96.63	96.75
HOG-BOW	94.50	<b>95.38</b>	91.88	92.88	94.00	93.63	93.13	93.88	94.75	93.63
MSMF-CLBP+ HOG-FV	99.38	99.50	<b>99.75</b>	99.25	99.38	99.25	99.25	99.63	99.25	99.50
ISS	85.63	85.63	85.75	85.63	86.75	<b>87.38</b>	86.00	86.00	86.38	85.63
HOG+CLBP	<b>98.50</b>	98.13	97.88	97.88	98.25	98.13	97.88	98.13	97.50	97.75

*Bold indicates the better performance for each method.*

**Table 3.** F1-measure (%) comparison with different features.

Different features	F <sub>1</sub> -measures of 10 random tests									
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
MSMF-CLBP	98.63	<b>98.88</b>	98.38	98.75	98.75	98.75	98.75	98.14	98.62	98.62
CLBP	94.85	94.90	96.12	95.66	95.76	95.88	96.11	96.10	<b>96.24</b>	95.41
MS-CLBP	97.76	97.88	97.01	96.99	97.89	<b>98.26</b>	96.89	97.76	97.50	96.50
HOG-FV	97.27	96.45	96.30	96.81	96.95	95.96	<b>97.28</b>	96.00	96.43	96.81
HOG	95.10	95.84	95.42	96.08	96.32	96.21	96.09	<b>96.77</b>	96.65	96.77
HOG-BOW	94.34	<b>95.23</b>	91.46	92.53	93.78	93.40	92.77	93.66	94.57	93.30
MSMF-CLBP+ HOG-FV	99.38	99.50	<b>99.75</b>	99.25	99.38	99.25	99.25	99.63	99.38	99.50

ISS	85.71	86.44	85.96	85.96	86.82	<b>87.55</b>	85.96	86.03	86.56	85.71
HOG+CLBP	<b>98.52</b>	98.16	97.91	97.91	98.28	98.16	97.92	98.15	97.55	97.79

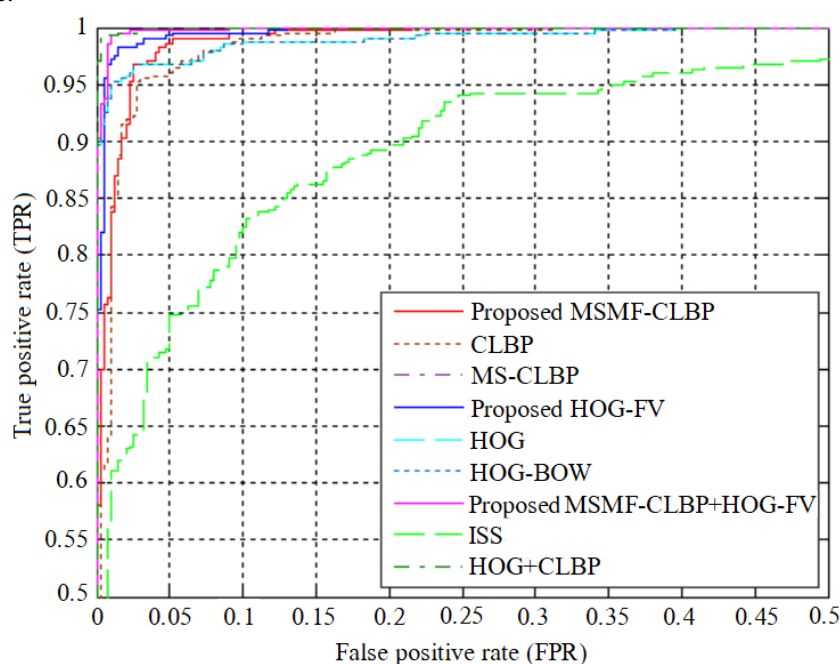
**Bold indicates the better performance for each method.**

**Table 4.** Average performance (%) comparison with different features.

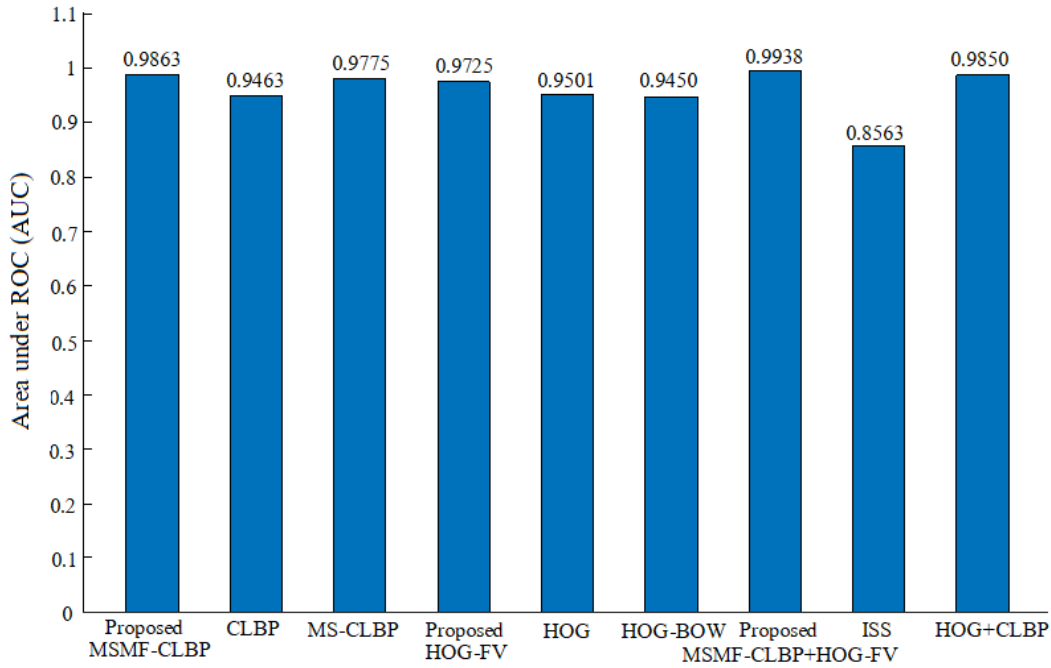
Different features	$\overline{AR}$	$\overline{F_1}$	SD
MSMF-CLBP	98.61	98.63	0.24
CLBP	95.56	95.70	0.53
MS-CLBP	97.44	97.45	0.56
HOG-FV	96.56	96.63	0.50
HOG	96.06	96.12	0.58
HOG-BOW	93.76	93.50	1.00
MSMF-CLBP+ HOG-FV	<b>99.41</b>	<b>99.43</b>	<b>0.18</b>
ISS	86.18	86.27	0.59
HOG+CLBP	98.00	98.03	0.28

**Bold indicates the best performance among all methods.**

**Fig. 4** shows the ROC curves of different features. As can be seen, the ROC curve of our proposed MSMF-CLBP+HOG-FV outperforms other features on target recognition performance for infrared images. **Fig. 4** also shows the ROC curves of the proposed MSMF-CLBP and the proposed HOG-FV are also higher than those of the commonly used features. Besides the ROC curve, the AUC is also computed for each method. The results are shown in **Fig. 5**, where we can see that our MSMF-CLBP+HOG-FV feature has the highest AUC value.



**Fig. 4.** Comparison results in terms of ROC curves with different features.



**Fig. 5.** Comparison results in terms of AUC with different features.

### 3.3 Evaluation of Improved Multi-kernel Transfer Learning Classifier

In this section, our proposed MK-TrAdaBoost classifier is evaluated by comparing it with the classic TrAdaBoost algorithm [26] as well as several state-of-the-art classifiers, including Random Forest classifier [34], Naive Bayes classifier [35], and Discriminant Analysis classifier [36], Adaboost classifier [37], KNN classifier [38]. Thereinto, Random Forest classifier, which contains multiple decision trees, can process high-dimension data and has good generalization performance, but it is prone to over-fitting in the classification process. Naive Bayesian classifier has stable effectiveness and is easy to implement, but it is only suitable for simple classification problems. Discriminant Analysis classifier can achieve classification results without feature selection, but it cannot handle the classification of high-dimension data. Adaboost classifier is a high-precision classifier, but its training process is time-consuming, and data imbalance may lead to a drop in classification accuracy. KNN classifier is very simple and efficient, but its performance is also subject to data imbalance problem. To compare the classification performance of various classification algorithms fairly, all of the above classifiers utilize the same feature, i.e., the proposed MSMF-CLBP+HOG-FV feature. The comparison results are shown in Table 5, Table 6 and Table 7.

From these three tables, it can be seen that, our proposed MK-TrAdaBoost achieves better classification performance with higher accuracy rate and F1-measure values and lower SD value, which proves that the recognition power is indeed enhanced by integrating the idea of the multi-kernel learning into the TrAdaBoost framework. In addition, compared with other state-of-the-art classifiers, the recognition performance of our classifier is more superior than the existing schemes. The results can be attributed to the multi-kernel learning and transfer learning ideas. Fig. 6 and Fig. 7 show more straightforward comparisons of ROC curves and AUC values. As can be seen, the proposed MK-TrAdaBoost classifier shows good performance in recognizing objects in infrared images.



**Table 5.** Accuracy rate (%) comparison with different classifiers.

Different classifiers	Accuracy rate of ten random tests									
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
MK-TrAdaBoost	99.38	99.50	<b>99.75</b>	99.25	99.38	99.25	99.25	99.63	99.25	99.50
TrAdaBoost	98.63	<b>98.88</b>	98.25	98.75	98.75	98.75	98.75	98.13	98.63	98.63
Random Forest	96.75	96.75	96.75	97.13	97.25	97.25	96.50	97.25	<b>97.63</b>	96.88
Naive Bayes	82.63	82.50	<b>83.25</b>	81.88	83.13	82.63	83.12	81.38	82.25	83.00
Discriminant Analysis	80.88	<b>81.50</b>	79.25	78.63	79.88	76.63	78.38	76.25	79.50	81.25
Adaboost	97.88	97.38	97.5	97.25	98.00	97.13	97.00	97.13	97.00	<b>98.13</b>
KNN	94.25	94.75	94.63	94.25	94.63	94.25	92.75	93.63	93.75	<b>94.88</b>

*Bold indicates the better performance for each method.*

**Table 6.** F1-measure (%) comparison with different classifiers.

Different classifiers	F <sub>1</sub> -measure of ten random tests									
	#1	#2	#3	#4	#5	#6	#7	#8	#9	#10
MK-TrAdaBoost	99.38	99.50	<b>99.75</b>	99.25	99.38	99.25	99.25	99.63	99.38	99.50
TrAdaBoost	98.63	<b>98.88</b>	98.38	98.75	98.75	98.75	98.75	98.14	98.62	98.62
Random Forest	96.83	96.82	96.82	97.18	97.30	97.30	96.59	97.30	<b>97.66</b>	96.95
Naive Bayes	85.10	85.11	<b>85.56</b>	84.59	85.44	85.07	85.47	84.27	84.83	85.47
Discriminant Analysis	80.85	<b>81.55</b>	78.39	78.81	79.44	75.87	78.07	76.01	79.45	81.48
Adaboost	97.89	97.40	97.52	97.28	98.02	97.17	97.06	97.13	97.00	<b>98.14</b>
KNN	94.31	94.87	94.65	94.28	94.65	94.31	92.77	93.54	93.75	<b>94.89</b>

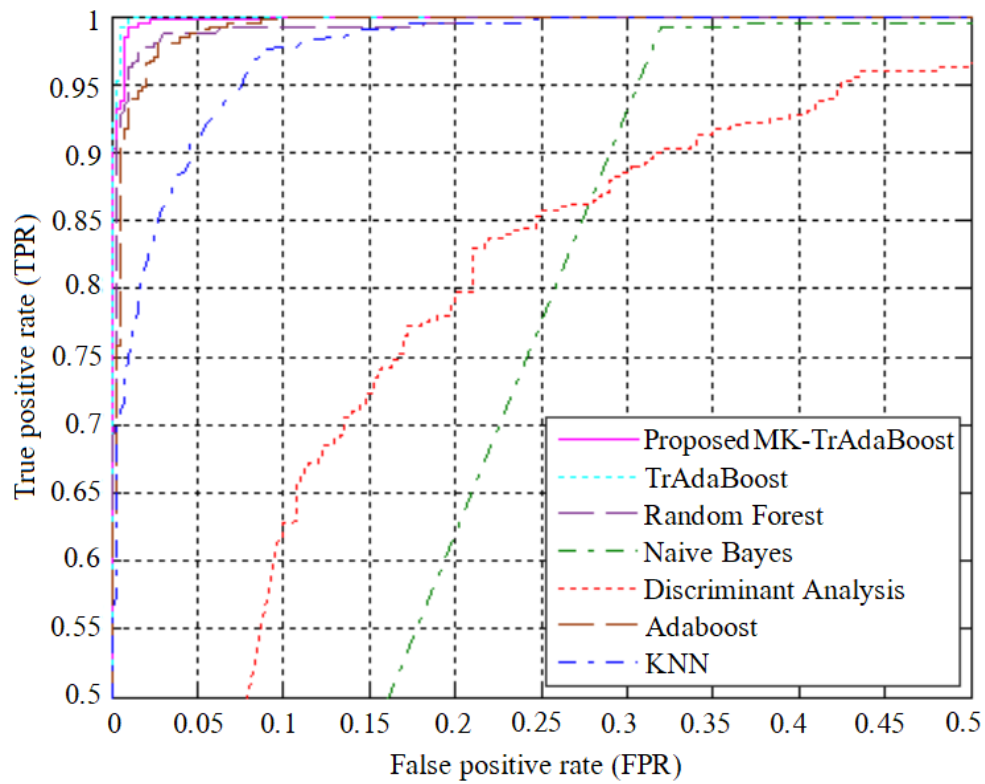
*Bold indicates the better performance for each method.*

**Table 7.** Average performance (%) comparison with different classifiers.

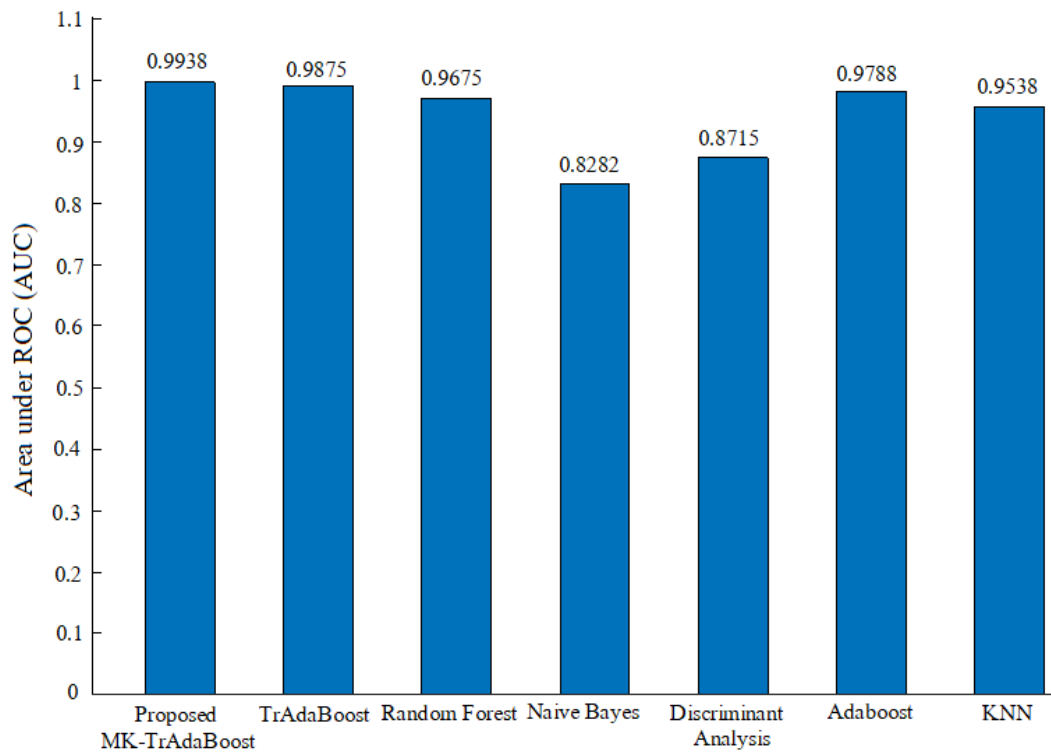
Different classifiers	$\overline{AR}$	$\overline{F_1}$	SD
MK-TrAdaBoost	<b>99.41</b>	<b>99.43</b>	<b>0.18</b>
TrAdaBoost	99.21	99.23	0.24
Random Forest	97.01	97.08	0.34
Naive Bayes	82.58	85.09	0.60
Discriminant Analysis	79.21	78.99	1.80
Adaboost	97.44	97.46	0.42
KNN	94.18	94.20	0.65

*Bold indicates the best performance among all methods.*





**Fig. 6.** Comparison results in terms of ROC curves with different classifiers.



**Fig. 7.** Comparison results in terms of AUC with different classifiers.

## 5. Conclusion

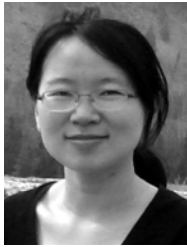
This paper introduces a recognition method based on heterogeneous features extraction and multi-kernel transfer learning classifier. The proposed method is applied to pedestrian target recognition in infrared images. Unlike the preceding works, in which a single kind of low-level features or homogeneous features are extracted, this paper proposes to extract high-level heterogeneous features. Second, an improved classification algorithm via multi-kernel learning and transfer learning has been developed to achieve target classification. To verify the proposed method, multiple comparative approaches have been studied, and the comparison results demonstrate that our method consistently exceeds the competitors. An intriguing question for future work is to apply the proposed method to other target recognition tasks, fully tapping the potential of the algorithm in target recognition. In addition, more algorithms, such as deep learning based feature extraction methods and classifiers, will be studied and used for comparison with the proposed method in the near future, so as to further illustrate the advantages and disadvantages of our algorithm for the infrared target recognition task.

## References

- [1] Kushwaha A K S and Srivastava R, "Multiview human activity recognition system based on spatiotemporal template for video surveillance system," *Journal of Electronic Imaging*, vol. 24, no. 5, pp. 051004, October, 2015. [Article \(CrossRef Link\)](#).
- [2] Kim D S, Kim M, Kim B S and Lee K H, "Histograms of local intensity differences for pedestrian classification in far-infrared images," *Electronics Letters*, vol. 49, no. 4, pp. 258-260, February, 2013. [Article \(CrossRef Link\)](#).
- [3] Lee Y S, Chan Y M, Fu L C and Hsiao P Y, "Near-Infrared-Based Nighttime Pedestrian Detection Using Grouped Part Models," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 1929-1940, August, 2015. [Article \(CrossRef Link\)](#).
- [4] Sun J, Fan G, Yu L and Wu, "Concave-convex local binary features for automatic target recognition in infrared imagery," *EURASIP Journal on Image and Video Processing*, vol. 2014, no. 1, pp. 23, February, 2014. [Article \(CrossRef Link\)](#).
- [5] Miron A, Besbes B, Rogozan A and Ainouz S, "Intensity self-similarity features for pedestrian detection in Far-Infrared images," in *Proc. of IEEE Conf. on Intelligent Vehicles Symposium*, pp. 1120-1125, June 3-7, 2012. [Article \(CrossRef Link\)](#).
- [6] Hurney P, Waldron P, Morgan F, Jones E and Glavin M, "Night-time pedestrian classification with histograms of oriented gradients-local binary patterns vectors," *IET Intelligent Transport Systems*, vol. 9, no. 1, pp. 75-85, January, 2014. [Article \(CrossRef Link\)](#).
- [7] Ragb H K and Asari V K, "Multi-feature fusion and PCA based approach for efficient human detection," in *Proc. of IEEE Conf. on Applied Imagery Pattern Recognition Workshop*, pp. 1-6, October 18-20, 2016. [Article \(CrossRef Link\)](#).
- [8] Bi Y, Lv M, Wei Y, Guan N and Yi W, "Multi-feature fusion for thermal face recognition," *Infrared Physics & Technology*, vol. 77, pp. 366-374, July, 2016. [Article \(CrossRef Link\)](#).
- [9] Kachach R and Plaza J M C, "Hybrid three-dimensional and support vector machine approach for automatic vehicle tracking and classification using a single camera," *Journal of Electronic Imaging*, vol. 25, no. 3, pp. 033021, June, 2016. [Article \(CrossRef Link\)](#).
- [10] Wu Y, Cheng Y, Zhao Y and Gao S, "Detection of infrared targets based on Adaboost by feature extraction using KPCA," *Infrared & Laser Engineering*, vol. 40, no. 2, pp. 338-343, February, 2011. [Article \(CrossRef Link\)](#).
- [11] Dawood H, Shabbir S, Dawood H and Majeed N, "Sparsely encoded distinctive visual features for object recognition," *Journal of Electronic Imaging*, vol. 27, no. 6, pp. 063035, December, 2018. [Article \(CrossRef Link\)](#).

- [12] Wang X, Shen S, Ning C, Zhang Y and Lv G, "Robust object tracking based on local discriminative sparse representation," *Journal of the Optical Society of America A*, vol. 34, no. 4, pp. 533-544, March, 2017. [Article \(CrossRef Link\)](#).
- [13] Pan S J and Yang Q, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no.10, pp. 1345-1359, October, 2010. [Article \(CrossRef Link\)](#).
- [14] Li X, Zhang L, Du B, Zhang L and Shi Q, "Iterative reweighting heterogeneous transfer learning framework for supervised remote sensing image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 10, no. 5, pp. 2022-2035, May, 2017. [Article \(CrossRef Link\)](#).
- [15] Malmgren-Hansen D, Kusk A, Dall J, Nielsen A A, Engholm R and Skriver H, "Improving SAR Automatic Target Recognition Models With Transfer Learning From Simulated Data," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 9, pp. 1484-1488, September, 2017. [Article \(CrossRef Link\)](#).
- [16] Ren C X, Dai D Q, Huang K K and Lai Z R, "Transfer Learning of Structured Representation for Face Recognition," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5440-5454, December, 2014. [Article \(CrossRef Link\)](#).
- [17] Guo Z, Zhang L and Zhang D, "A completed modeling of local binary pattern operator for texture classification," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1657-1663, June, 2010. [Article \(CrossRef Link\)](#).
- [18] Chen C, Zhang B, Su H, Li W and Wang L, "Land-use scene classification using multi-scale completed local binary patterns," *Signal, image and video processing*, vol. 10, no. 4, pp. 745-752, July, 2016. [Article \(CrossRef Link\)](#).
- [19] Felsberg M and Sommer G, "The monogenic signal," *IEEE Transactions on Signal Processing*, vol. 49, no. 12, pp. 3136-3144, December, 2001. [Article \(CrossRef Link\)](#).
- [20] Ning C, Liu W and Wang X, "Infrared Object Recognition Based on Monogenic Features and Multiple Kernel Learning," in *Proc. of IEEE Conf. on Image, Vision and Computing*, pp. 204-208, June 27-29, 2018. [Article \(CrossRef Link\)](#).
- [21] Dong G , Kuang G , Wang N, Zhao L and Lu J, "SAR Target Recognition via Joint Sparse Representation of Monogenic Signal," *IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing*, vol. 8, no. 7, pp. 3316-3328, July, 2015. [Article \(CrossRef Link\)](#).
- [22] Ning C, Liu W, Zhang G and Yin J, "Enhanced synthetic aperture radar automatic target recognition method based on novel features," *Applied Optics*, vol. 55, no. 31, pp. 8893-8904, November, 2016. [Article \(CrossRef Link\)](#).
- [23] Khan M N A, Fan G, Heisterkamp D R and Yu L, "Automatic target recognition in infrared imagery using dense HOG features and relevance grouping of vocabulary," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition Workshops*, pp. 293-298, June 23-28, 2014. [Article \(CrossRef Link\)](#).
- [24] Sánchez J, Perronnin F, Mensink T and Verbeek J, "Image classification with the fisher vector: Theory and practice," *International journal of computer vision*, vol. 105, no. 3, pp. 222-245, June, 2013. [Article \(CrossRef Link\)](#).
- [25] Kawano Y and Yanai K, "Rapid mobile object recognition using fisher vector," in *Proc. of IEEE Asian Conf. on Pattern Recognition*, pp. 476-480, November, 2013. [Article \(CrossRef Link\)](#).
- [26] Dai W, Yang Q, Xue G R and Yu Y, "Boosting for transfer learning," in *Proc. of ACM International Conf. on Machine Learning*, pp. 193-200, June, 2007. [Article \(CrossRef Link\)](#).
- [27] Gu Y, Wang Q, Jia X and Benediktsson J, "A novel MKL model of integrating LiDAR data and MSI for urban area classification," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 10, pp. 5312-5326, May, 2015. [Article \(CrossRef Link\)](#).
- [28] Wang X, Xiong X, Ning C and Shi A, "Integration of heterogeneous features for remote sensing scene classification," *Journal of Applied Remote Sensing*, vol. 12, no. 1, pp. 015023, March, 2018. [Article \(CrossRef Link\)](#).
- [29] Khellal A, Ma H and Fei Q, "Pedestrian Classification and Detection in Far Infrared Images," in *Proc. of International Conf. on Intelligent Robotics and Applications*, pp. 511-522, August, 2015. [Article \(CrossRef Link\)](#).

- [30] Dalal N, "Finding people in images and videos," *Institute National Polytechnique de Grenoble-INPG*, July, 2006. [Article \(CrossRef Link\)](#).
- [31] Wang X, Shen S, Ning C, Huang F and Gao H, "Multi-class remote sensing object recognition based on discriminative sparse representation," *Applied optics*, vol. 55, no. 6, pp. 1381-1394, February, 2016. [Article \(CrossRef Link\)](#).
- [32] Wang X, Zhang Y and Ning C, "A novel visual saliency detection method for infrared video sequences," *Infrared Physics & Technology*, vol. 87, pp. 91-103, December, 2017. [Article \(CrossRef Link\)](#).
- [33] Hassan M A, Pardiansyah I, Malik A S, Faye I and Rasheed W, "Enhanced people counting system based head-shoulder detection in dense crowd scenario," in *Proc. of IEEE Conf. on Intelligent and Advanced Systems*, pp. 1-6, August, 2016. [Article \(CrossRef Link\)](#).
- [34] Lee E J, Ko B C and Nam J Y, "Recognizing pedestrian's unsafe behaviors in far-infrared imagery at night," *Infrared Physics & Technology*, vol. 76, pp. 261-270, May, 2016. [Article \(CrossRef Link\)](#).
- [35] Bo Y, Lei Y and Bei Y, "Distributed Multi-Human Location Algorithm Using Naive Bayes Classifier for a Binary Pyroelectric Infrared Sensor Tracking System," *IEEE Sensors Journal*, vol. 16, no. 1, pp. 216-223, January, 2016. [Article \(CrossRef Link\)](#).
- [36] Zhan W, Ruan Q and An G, "Facial expression recognition using sparse local Fisher discriminant analysis," *Neurocomputing*, vol. 174, pp. 756-766, January, 2016. [Article \(CrossRef Link\)](#).
- [37] Peng D, Chen Y and Yue H, "Remote-sensing imagery classification using multiple classification algorithm-based AdaBoost," *International Journal of Remote Sensing*, vol. 39, no. 3, pp. 619-639, October, 2017. [Article \(CrossRef Link\)](#).
- [38] Sánchez A S, Iglesias-Rodríguez F J, Fernández P R and Juez F J D C, "Applying the K-nearest neighbor technique to the classification of workers according to their risk of suffering musculoskeletal disorders," *International Journal of Industrial Ergonomics*, vol. 52, pp. 92-99, March, 2016. [Article \(CrossRef Link\)](#).



**Xin Wang** received the Ph.D. degree in Computer Application Technology from Nanjing University of Science and Technology, Nanjing, China, in 2010. She is currently an Associate Professor with the College of Computer and Information, Hohai University, Nanjing, China. She has published more than 50 papers in journals and referred conferences. Her current research interests include image processing, computer vision, and pattern recognition.



**Xin Zhang** received the B.S. degree in Communication Engineering from Hohai University, Nanjing, China, in 2017. Now He is working toward the M.S. degree in the College of Computer and Information, Hohai University. His current research interests include image processing, target detection and recognition.



**Chen Ning** received the M.S. degree in Signal and Information Processing from University of Science and Technology of China, Hefei, China, in 2003. He is currently working in the School of Physics and Technology, Nanjing Normal University, Nanjing, China. His current research interests include signal and image processing, and computer vision.