Target Detection and Pedestrian Recognition in Infrared Images

Jiabao Wang*
Institute of Command Automation, PLA Univ. of Sci. and Tech., Nanjing, 210007, China
Email: boa.plaust@gmail.com

Yafei Zhang, Jianjiang Lu and Yang Li
Institute of Command Automation, PLA Univ. of Sci. and Tech., Nanjing, 210007, China
Email: Email: miipl606@163.com

Abstract—By improving the local contrast between targets and background in the static infrared images, a simple and effective background model is proposed to detect targets. At the same time, a novel learning algorithm is presented for training a discriminatively trained, part-based model with only positives images, for pedestrian recognition. The background models are constructed based on the static infrared images by morphological operations. Meanwhile, the learning algorithm is based on the ramp loss function, which can filter out the false negatives from the collected negative examples. It has a great advantage on training the deformable part models with latent variables when the dataset has a large number of noisy examples. Experiments manifest that our background model can achieve a high precision in target detection and the discriminative part model trained by the proposed learning approach can recognize the targets well and truly, with the help of target detection.

Index Terms—infrared images, target detection, pedestrian recognition, ramp loss, stochastic gradient descent

I. INTRODUCTION

Recently, target detection and pedestrian recognition in infrared images have become one of the research hottest topics due to the widely used infrared sensors in public surveillance. It has many significant applications, such as military target recognition and tracking [1], public traffic surveillance and guidance [2], detection and alarming of the high dangerous areas [3], and also achieved a great development. However, there are still many problems existed in target detection and pedestrian recognition in infrared images. Firstly, infrared images are static images, among which there are no associations over time like the frames in videos. The traditional target detection methods based on segmentation are hard to separate targets from the background, especially when the features of targets are similar to the disturbance things, such as a pedestrian is hard to distinguish from a vertical strip lamp in infrared images. Secondly, the Signal to Noise Ratio (SNR) is low in infrared images and then the edges of targets and local areas of images are often blurred. It is difficult to segment targets from the background completely and to recognize their categories. Thirdly, targets are prone to occluded or truncated by other things in surveillance.

Target detection methods can be mainly divided into single image-based detection methods [4], [5], [6], [7] and sequential images-based detection methods [8], [9], [10], [11]. The former usually segments targets from the background based on the features of the infrared images such as color, texture, shape and so on. They partition the image into perceptually similar areas, two of the famous ones are the mean shift approach [6] and the graph cuts approach [7]. Mean shift segmentation approach requires fine tuning of various parameters to obtain better results, while graph cuts segmentation requires a large number of memory and computing [12]. By comparison, sequential images-based detection is based on the connection among the two or more consecutive frames, such as background-differencing [9], frame-differencing [1], [10] motion history image [11], and optical flow [13], [14]. However, these approaches have a significant defect that SNR in infrared images must be high for detection.

The basic process of the target recognition is: firstly, to extract the features of targets in the training datasets, and then to train the target classifiers based on the extracted features, finally, to recognize the categories of the targets by the trained classifiers. In infrared target recognition, principle component analysis and independent component analysis [15], [16] are two famous techniques. However, both of them deal with the pixels of the target directly and indirectly. They have to face high dimension features of the targets and a large number of computations. Although many simple features have been proposed to describe the targets, such as length-width ratio, standard deviation and variance, contrast, invariant moment, they are too simple to describe or represent the targets, and usually give a high false recognition rate.

To overcome these problems in target detection and pedestrian recognition, in this paper, we proposed a novel target detection approach based on background model, through which the SNR can be enhanced and the anti-
jamming ability can be improved. At the same time, we proposed a robust learning method for object recognition based on the Ramp Loss-based Support Vector Machines (SVMs) [17] and the extended Histogram of Orientation Gradient (HOG) features [18], which can describe the shape and the pose of targets and achieve a great success in object detection.

The remainder of this paper is structured as follows: in Section 2, we present the overall framework of our target detection and pedestrian recognition; in Section 3, the background model-based target detection is discussed; in Section 4, the robust learning algorithm of the pedestrian recognition is given; in Section 5, we experiment on the famous dataset: OTCVBS thermal pedestrian dataset. Finally, we conclude in Section 6.

II. FRAMEWORK

Fig. 1 shows the framework of our proposed target detection and pedestrian recognition. It includes two main modules: target detection and pedestrian recognition. The former is divided into four main steps: background model construction, contrast enhancement, foreground detection and target detection. The latter includes classifier training, pedestrian recognition and target relocation. Both of the two modules are connected to each other. The results of target detection provide the positions and sizes of targets and reversely pedestrian recognition gives more accurate positions and sizes of the targets. The novelties of this framework are concluded as follows: 1) Good expansion of the modules, both of them not only can be served for detecting targets and recognizing pedestrians in the static images, but also can be applied in the infrared videos or surveillance in real-time. 2) Low computing complexity, we detect targets based on only one frame and recognize pedestrians on the regions of target detection, which can filter out a large number of useless regions in the infrared images, so the cost of the overall computing is low.

III. TARGET DETECTION

A. Background Model

The problems of target detection we confronted in the infrared images can be concluded as follows: 1) when the difference of intensities between target and background is small, the detected targets may be divided into multi-parts or only one local area when the Gaussian-based model is used; 2) the distribution of the intensities in an infrared image usually does not follow a Gaussian distribution and the distributions may be different in different time, scene or temperature; 3) the means and variances of intensities in difference infrared images may be different and so the parameters have to be tuned by human labors.

In this paper, a simple and effective background model is presented for target detection. It improves the precision of target detection by enhancing the contrast between the targets and the background. The construction processes of the background model are:

**Step 1:** Initialize the filter with a kernel of fixed width \( c = 0.5 \times \max(w, h) \), where \( w \) and \( h \) are the average width and height of the targets in training datasets, and then execute a dilation operation on the given infrared images to remove the foreground pixels of targets by the formulation:

\[
D(i, j) = \max_{s,t} I(i + s, j + t),
\]

where \( I \) is the given infrared image. The width \( w \) and height \( h \) of the targets in a fixed surveillance scene can be estimated in prior.

**Step 2:** Erode to the result of the Step 1 with a kernel of the same width. This step is to recover the foreground areas

\[
E(i, j) = \min_{s,t} D(i + s, j + t).
\]

Based on above two steps, we can get the background model. Note that the width of the kernel is a well-chosen value, which is to decide if the target foreground pixels can be removed or not. In this work, we choose a fixed width in prior. The main process of detecting targets is, firstly, to remove the background from the given images by subtracting the computed background model, and then extract the foreground targets. Although the model and the process are simple, the effect is remarkable. Fig. 2(a) is a given infrared image, Fig. 2(b) is the corresponding foreground detecting result, and Fig. 2(c) is the result of target detection. Accordingly, Fig. 2(d) is the image with removing the background model, Fig. 2(e) and (f) are the results of foreground and target detection respectively. Apparently, our background model-based approach has a better detection results and precision.

Why our proposed background model achieves such a good result? Because it can reduce the mean and variance...
is the variance of $v_y$. We have $Y = \{x_i, y_i\}$, the output $X = \{x_i\}$, and $h = \{h_i\}$, that is $\langle \Psi \rangle$. Let $\hat{w}$ be a joint feature map. Then parameters $\mu$ and $\sigma$ are computed.

**Step 3:** Estimate the background by Gaussian model with the confidence value $\hat{T} = 0.997$, that is, we can estimate the foreground $B$:

$$B(i, j) = \begin{cases} 0, & \text{if } v_y \in [\mu - 3\sigma, \mu + 3\sigma] \\ 1, & \text{otherwise} \end{cases}$$

**Step 4:** According to the foreground image, detect the target by connection components approach [19].

### IV. PEDESTRIAN RECOGNITION

In this work, the problem of pedestrian recognition is a binary classification between pedestrian and non-pedestrian. Recently, discriminative models achieve great progress and get widely applications due to the simplicity and modeling without any prior knowledge. Especially, Felzenszwalb et al. proposed the discriminatively trained part model to improve the non-rigid object detection [18]. Their method achieved great precisions in non-rigid and occluded objects. As a result, in this paper, we adopt this discriminative latent part model to recognize pedestrians in infrared images. It reflects the distinct between the targets and background and can recognize and locate the targets. However, it is difficult to collect training datasets for learning pedestrian classifier, because images without non-pedestrian are not given directly. So we presented a new data collection method. It selects negative examples from positive images randomly and so there is many false negative examples produced. To deal with this problem, a Ramp Loss-based support vector machine is adopted to learn the filter by suppressing the false negative examples.

#### A. Problem

In general, for learning pedestrian classifiers, we need to learn a decision function $y = f(x), x \in X, y \in Y$ by which returns the label $y$ of a specific object in a given image $x$. The Structural SVM with latent variables [20] is adopted to specify the refinement of the ground-truth bounding boxes with the input variables $x$, the output variables $y$ and the auxiliary latent variables $h \in H$. We define the function $f$ as $f(x; w) = \hat{y}(w)$ where

$$\hat{y}(w) = \arg \max_{(x, y, h)} \langle \Psi(x, y, h) \rangle,$$  \tag{6}$$

$\Psi(x, y, h)$ is a joint feature map.

Given the training dataset $\{(x_i, y_i)\}_{i=1}^n$, the parameter filter $w$ can be learned by minimizing the following regularized empirical risk:

$$J(w) = \frac{1}{2} ||w||^2 + C \sum_{i=1}^n \Delta(y_i, \hat{y}_i(w), \hat{h}_i(w)),$$  \tag{7}$$

where $C$ is the penalty factor and $\Delta(y_i, \hat{y}_i(w), \hat{h}_i(w))$ is user-supplied loss function that encodes the cost of an incorrect prediction.
Actually, suppose the \(i\)th positive example located in the positive image \(x_i\) has the bounding box \(h_i\) and its label \(y_i = 1\). If we have only positive images, how do we collect the negative examples without given the negative images? Here, we present the new collection method. It selects both positive and negative examples from positive images, and the negative examples are randomly selected without any filtration. This collection strategy has a big problem that there are many false negatives among the negative examples. In this paper, we solve this problem by introducing Ramp Loss function \(\Delta = R_s(z)\), which has the formulation:

\[
R_s(z) = H_s(z) - H_1(z) = \min(s - z, \max(0, 1 - z)),
\]

(8)

where \(H_s(z) = \max(0, s - z)\) represents a class of loss functions. When \(s = 1\), \(H_1(z) = \max(0, 1 - z)\) represents the classical Hinge Loss. Equation (8) represents a class of loss functions decided by the parameter \(s\), \(z = y_i f(x_i)\) represents the score of the \(i\)th example. The Ramp Loss function can suppress the influence of examples with score \(z < s\) by not converting them into Support Vectors (SVs) and only allows examples with score \(z \in [s, 1]\) to be SVs. It implies that the Ramp Loss function has the function of prohibiting the false negative examples (noise or outliers) becoming SVs and affecting the hyper-plane of the classifier. Therefore, Ramp Loss-based SVMs can reduce the number of SVs and improve the generalization performance efficiently.

**B. Optimization**

Minimizing the regularized risk \(J(w)\) as defined by (7) is difficult because the loss function depends on the parameter \(w\) through the latent variables \(\hat{h}_i(w)\). To overcome this problem, it is possible to optimize an upper bound [21]:

\[
\Delta(y_i, \hat{y}_i(w), \hat{h}_i(w)) \leq Q(y, h),
\]

(9)

\[
Q(y, h) = \max_{y \in \hat{y}, h \in \hat{h}} \Delta(y, y, h)[1 + \langle w, \Psi(x, y, h) \rangle - \langle w, \Psi(x, y_i, h_i(w)) \rangle],
\]

(10)

where \(\hat{h}_i^*(w) = \arg \max_{h \in \hat{h}} \langle w, \Psi(x_i, y_i, h) \rangle\) completes the label \((y_i, \hat{h}_i^*(w))\) of the instance \(x_i\). By the above upper bound on the risk, the latent Structural SVMs [20] can be minimized based on CCCP and proceeds iteratively, and in each iteration there are two steps:

1. Imputing the latent variables \(\hat{h}_i^*(w)\) which correspond to approximating the concave function part by a linear upper bound;
2. Updating the new parameter \(w^{i+1}\) using the completed latent variables \(\hat{h}_i^*(w)\) as if they were completely observed, that is, a traditional Structural SVM learning problem is need to solve.

Above procedure has already been investigated by Yu and Joachims [20] and been improved by Kumar et al. [22]. The overall learning procedure is depicted in **Algorithm 1**.

**Algorithm 1**: The Overall Learning Procedure

1. Initialize \(w^0\) and \(\tau = 1\);
2. Compute: \(w_\tau = \arg \min_{w} \left( J_{w\tau}(w) + J_{w\tau}^*(w_{\tau-1}) \cdot w \right)\);
3. If \(| w_{\tau-1} - w_\tau | < \varepsilon\), stop; otherwise, set \(\tau = \tau + 1\) and go to Step 2.

**C. Sub-optimization procedure**

However, note that if we introduce the non-convex Ramp Loss function, the convex optimization problem turns into a non-convex optimization problem, which is a more complicated problem and the sub-optimization (Step 2 in **Algorithm 1**) is no longer to solve a structural SVMs [23], [24]. Fortunately, the non-convex objective function \(J^*(w)\) can be transformed into:

\[
J^*(w) = \frac{1}{2} \| w \|^2 + C \sum_{i \in \Omega} R_i(y_i f(x_i))
\]

(11)

\[
= \frac{1}{2} \| w \|^2 + C \sum_{i \in \Omega} H_i(y_i f(x_i)) - C \sum_{i \in \Omega} H_i(y_i f(x_i))
\]

that is the sum of a convex function \(J_{w\tau}(w)\) and a concave function \(J_{w\tau}^*(w)\). This minimum problem of this formulation can be solved by CCCP [25].

According to CCCP, optimization problem (11) can be solved by **Algorithm 2**:

**Algorithm 2**: The CCCP algorithm

1. Initialize: \(w_0\) and \(\tau = 1\);
2. Compute: \(w_\tau = \arg \min_{w} \left( J_{w\tau}(w) + J_{w\tau}^*(w_{\tau-1}) \cdot w \right)\);
3. If \(| w_{\tau-1} - w_\tau | < \varepsilon\), stop; otherwise, set \(\tau = \tau + 1\) and go to Step 2.

It provides a basic procedure for the non-convex optimization problem, Step 2 is executed to reduce the objective and finally this procedure can converge to a local minimum [25]. Based on this procedure, the traditional methods solve above non-convex optimization problem (11) by sequential minimal optimization algorithm [26]. However, these methods are extremely time-consuming when dataset is large-scale. Recently, the Stochastic Gradient Descend (SGD) algorithm [27], [28], which is able to fast obtain an approximate solution for a convex optimization problem, becomes popular. Therefore, in this paper we use the SGD algorithm to solve iteration Step 2. The SGD algorithm does not decrease the objective in each iteration but it still has an asymptotically convergence in the case of large-scale learning problems.
1) The SGD algorithm

The soft SVM problems [29] can be represented as:

$$
\min_w J(w) = \frac{1}{2} \| w \|^2 + C \sum_{i=1}^n \max(0, 1 - y_i f(x_i)) .
$$

(12)

The traditional gradient descent algorithms update the parameter $w$ based on the whole dataset in each iteration. However, SGD algorithm updates the parameter with only one random selected example. Since objective function (12) is not differentiable everywhere, we update the parameter $w$ by sub-gradient [30] as following:

$$
w_t = \begin{cases} 
    w_{t-1} - \eta_t (w_{t-1} + C y_i x_i), & \text{if } y_i f(x_i) < 1 \\
    w_{t-1} - \eta_t w_{t-1}, & \text{otherwise}
\end{cases},
$$

(13)

where $(x_i, y_i)$ is a randomly selected example at every iteration; $\eta_t = C / (t + t_0)$ is the learning rate; and $t_0$ is a constant, which is chose heuristically to keep the parameter $w$ not too big [31]. The SGD algorithm is illustrated in Algorithm 3:

Algorithm 3: The SGD algorithm

1. Initialize: $w_0$, $t = 1$;
2. Choose an example $(x_i, y_i)$ randomly, update
   
   $$
w_t = w_{t-1} - \eta_t \frac{\partial g}{\partial w} (w_{t-1}), \quad \text{where } \eta_t = C / (t + t_0) \quad \text{and} \quad g(w) = \frac{1}{2} \| w \|^2 + CH_i (y_i f_i(x_i));
   $$
3. If $\| w_{t-1} - w_t \| < \varepsilon$ or reach the maximal iteration, stop; otherwise, set $t = t + 1$ and go to Step 2.

2) The CCCP-SGD algorithm

Based on Algorithm 2 and Algorithm 3, we solve the non-convex optimization problem (11) from primal, and the methods from dual are referred to [17], [32]. The Step 2 in Algorithm 2 is to compute parameter $w_t$ by optimizing the convex sub-problems. According to CCCP, the differentiable of concave part $J_w^\prime (w)$ is:

$$
\frac{\partial J_w^\prime (w)}{\partial f_i (x_i)} = \begin{cases} 
    C y_i, & \text{if } y_i f_i (x_i) < s \\
    0, & \text{otherwise}
\end{cases}.
$$

(14)

If $y_i f_i (x_i) < s$, the $g(w)$ in the Step 2 of Algorithm 2 can be represented as:

$$
g(w) = \frac{1}{2} \| w \|^2 + CH_i (y_i f_i (x_i)) + C y_i f_i (x_i) \cdot w,
$$

(15)

Otherwise,

$$
g(w) = \frac{1}{2} \| w \|^2 + CH_i (y_i f_i (x_i)).
$$

(16)

Equations (15) and (16) can be adopted to update the parameter $w_t$. As a result, those convex sub-problems can be solved by SGD algorithm, with a little difference in objective function in each iteration.

The basic procedure of our learning algorithm for the non-convex linear SVMs based on SGD in this work is summarized in Algorithm 4. SGD algorithm can converge to the minimal expected risk, but the convergence speed is slower than that of the traditional gradient descent algorithms due to the influence of the noisy data [32]. In machine learning problems, the empirical risk is only an approximate of the expected risk, and the real interest of us is the latter. Therefore, the learning algorithm based on stochastic optimization converges to the minimal expected risk and reaches the objective in the end.

Algorithm 4: The CCCP-SGD algorithm

1. Initialize: $\hat{w}_0$, $\tau = 1$;
2. Compute $\hat{w}_\tau$:
   
   (a) Initialize: $w_{\tau-1} = \hat{w}_{\tau-1}$, $t = 1$;
   
   (b) Choose an example $(x_i, y_i)$ randomly, update
   
   $$
w_t = w_{t-1} - \eta_t \frac{\partial g}{\partial w} (w_{t-1}), \quad \text{where } \eta_t = C / (t + t_0) \quad \text{and} \quad g(w) \text{ is defined by (15) and (16)};
   $$
3. If $\| w_{\tau-1} - w_t \| < \varepsilon$ or reach the maximal iterations, set $\hat{w}_\tau = w_t$ and go to Step 3; otherwise, set $\tau = \tau + 1$ and go to Step (b).

V. EXPERIMENTS

Our experiments are performed on OTCVBS thermal pedestrian dataset, which has 10 subsets from 00001 to 00010, including 284 pictures and 984 pedestrian labeled examples. Several kinds of conditions appeared in the surveillance scene: rainy and windy weather, foggy and sunny and so on showed in Fig. 4(a). Besides, pedestrians have many complicated gestures and appearances: running, walking, standing still, and with backpack, umbrella, raincoat and so on. So many problems make it difficult to detect target and recognize pedestrian.

The length and width of the targets in our experiment are about 25×20 pixels, so we use square filter kernel with the width $\varepsilon = 15$. We expand the regions of target detection by the same size for recognizing pedestrian in a large area. The aim is to make sure that the targets are
covered in detected areas even if the results of detection just are local part of targets. Besides, we zoom in the expanded regions by 4 times because the original areas are too small to extract HOG features effectively.

A. Evaluation

For evaluating the performance of our approach, some criterions for target detection and pedestrian recognition are defined. Suppose the number of targets in the whole dataset is $N$, the number of detected targets is $N_{det}$ and the number of recognized targets is $N_{rec}$. The number of

![Examples of target detection](image1)

![Examples of pedestrian recognition](image2)

Figure 5. Examples of target detection and pedestrian recognition
correct detection is \( N_{det}^C \), the number of false detection is \( N_{det}^F \), the number of leaked detection is \( N_{det}^L \); the number of correct recognition is \( N_{rec}^C \), the number of false recognition is \( N_{rec}^F \), the number of leaked recognition is \( N_{rec}^L \). We define the evaluation criterion as follows:

Target detection evaluation criterion:

\[
\text{Correct detection rate (CDR)} = \frac{N_{det}^C}{N_{det}}; \\
\text{False detection rate (FDR)} = \frac{N_{det}^F}{N_{det}}; \\
\text{Leaked detection rate (LDR)} = \frac{N_{det}^L}{N}.
\]

Pedestrian recognition evaluation criterion:

\[
\text{Correct recognition rate (CRR)} = \frac{N_{rec}^C}{N_{rec}}; \\
\text{False recognition rate (FRR)} = \frac{N_{rec}^F}{N_{rec}}; \\
\text{Leaked recognition rate (LRR)} = \frac{N_{rec}^L}{N}.
\]

### B. Results

According to evaluation criterions in 5.1, experimental results are showed in Table I.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Target Detection</th>
<th>Pedestrian Recognition</th>
</tr>
</thead>
<tbody>
<tr>
<td>CDR FDR LDR</td>
<td>CRR FRR LRR</td>
<td></td>
</tr>
<tr>
<td>00001</td>
<td>0.9419 0.0581 1.0000</td>
<td>1.0000 0.0000 0.3000</td>
</tr>
<tr>
<td>00002</td>
<td>0.6691 0.3309 0.9000</td>
<td>1.0000 0.0000 0.0800</td>
</tr>
<tr>
<td>00003</td>
<td>0.1368 0.8632 0.8750</td>
<td>0.8182 0.1818 0.9135</td>
</tr>
<tr>
<td>00004</td>
<td>0.8583 0.1417 0.0180</td>
<td>1.0000 0.0000 0.0360</td>
</tr>
<tr>
<td>00005</td>
<td>0.4973 0.5027 0.0792</td>
<td>1.0000 0.0000 0.0495</td>
</tr>
<tr>
<td>00006</td>
<td>0.7203 0.2797 0.1053</td>
<td>0.9756 0.0244 0.1579</td>
</tr>
<tr>
<td>00007</td>
<td>0.9032 0.0968 0.1340</td>
<td>1.0000 0.0000 0.2887</td>
</tr>
<tr>
<td>00008</td>
<td>0.7881 0.2119 0.0606</td>
<td>1.0000 0.0000 0.2626</td>
</tr>
<tr>
<td>00009</td>
<td>0.5000 0.5000 0.0000</td>
<td>1.0000 0.0000 0.0000</td>
</tr>
<tr>
<td>00010</td>
<td>0.4397 0.5603 0.4688</td>
<td>1.0000 0.0000 0.5567</td>
</tr>
<tr>
<td>Average</td>
<td>0.6455 0.3545 0.1931</td>
<td>0.9794 0.0206 0.2645</td>
</tr>
</tbody>
</table>

The CRR in experiment is 0.9794, which is enough to meet the demand in real-life. But the average LRR is 0.2645, which is a little too high. The reasons can be concluded as: 1) the existence of disturbance, such as the street lamp, vehicle, trunk and the edge of infrared images. 2) the noise at the edge of infrared images. 3) the appearance of targets change too much. So the lower LDR results in higher LRR.

### C. Comparison

In this additional part, the aim is to prove that CCCP-SGD algorithm has a strong ability to suppressing false negative examples and produces a strong classifier with high generalization performance in recognition.

Firstly, the classifiers learned by SVM-SGD proposed by Bottou et al. [27] and CCCP-SGD are evaluated on our collected positive and negative examples. As shown in Table II, CCCP-SGD has a larger training error, but it achieves a lower testing error for the collected training examples, which contains a lot of false negative examples. Contrastively, SVM-SGD is over fit in learning.

Secondly, the best bounding boxes are estimated for positive examples through Latent-SSVMs. Experimental results are show in Fig. 6. In Fig. 6, we use the evaluation standard of the PASCAL 2010 detection competition, the Precision-Recall (PR) curves and the Average Precision (AP). From the PR curves and APs, we can infer that Latent-SSVMs have a great advantage on object detection.

### VI. Conclusions

In this paper, we design a novel framework for infrared target detection and pedestrian recognition. And more, a new background model is presented for target detection and the discriminative latent part model is adopted for pedestrian recognition. It is able to detect targets and recognize pedestrians in real time and can be applied to.
many important applications, such as, scene surveillance and target reconnaissance. Besides, both target detection and pedestrian recognition can also be applied for online detecting and tracking pedestrians in many military areas or dangerous areas.

ACKNOWLEDGMENT

This work was supported in part by NSFC under Grant Nos. 61070173.

REFERENCES


Jiabao Wang received the Bachelor's degree in computer science and technology from PLA University of Science and Technology, Nanjing, China, in 2008. He is currently a Ph.D. candidate in PLA University of Science and Technology. His research interests include intelligent information processing and pattern recognition.

Yafei Zhang received the Ph.D. degree from Fudan University, Shanghai, China, in 1992. Now he is a professor with the Department of Computer Science and Engineering, PLA University of Science and Technology. His research interests include computational intelligence, pattern recognition, machine learning, natural language processing, and distributed system design and optimization.