Pattern Recognition 48 (2015) 1947–1960

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Robust pedestrian detection in thermal infrared imagery using a shape distribution histogram feature and modified sparse representation classification

Xinyue Zhao [a], Zaixing He [a,*], Shuyou Zhang [a], Dong Liang [b]

[a] The State Key Lab of Fluid Power and Mechatronic Systems, Zhejiang University, Hangzhou, China
[b] Department of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Yudao Street 29, Nanjing 210016, PR China

ARTICLE INFO

Article history:
Received 10 December 2013
Received in revised form
30 November 2014
Accepted 12 December 2014
Available online 24 December 2014

Keywords:
Shape distribution histogram (SDH) feature
Modified sparse representation classification (MSRC)
Pedestrian detection
Thermal infrared imagery

ABSTRACT

In this paper, a robust approach using a shape distribution histogram (SDH) feature and modified sparse representation classification (MSRC) for pedestrian detection in thermal infrared imagery is proposed. In this framework, the candidate regions that are more likely to contain the pedestrians are first detected based on the Contour Saliency Map. Then distances between random points on the thinned contour map of objects in the candidate regions are applied to acquire the SDH feature. SDH is a robust and discriminative feature which can precisely describe the pedestrian characteristics. Finally, a robust MSRC classifier which has high accuracy is used to recognize the true pedestrians. Experiments are conducted over the OSU thermal pedestrian database by comparing with other algorithms. The proposed method shows an excellent performance in detecting pedestrians.

1. Introduction

Pedestrian detection is an essential and significant task in the field of computer vision. It provides the fundamental information for many vision-based applications, such as visual surveillance and traffic monitoring. Recently, pedestrian detection in thermal imagery using infrared cameras has attracted more and more attention, since infrared images can eliminate the influences of color and illumination variations compared with visible light images. However, thermal imagery has its own unique challenges [1]. First, most thermal imageries have low image qualities due to the low SNR of thermal sensors. They cannot provide as much information as visible ones can about objects. Second, non-human objects and backgrounds always produce additional bright areas which disturbs the detection. Third, the thermal infrared imagery does not depend on lighting conditions but on temperature changes. Thus, the human body appears bright on cold days and turns to dark on hot summer days. This makes standard background-subtraction and template matching techniques ineffective to accurately detect pedestrians.

Nowadays, different algorithms have been proposed for pedestrian detection in thermal infrared imagery. Nanda et al. used probabilistic templates to capture the variations in human shape for pedestrian detection [2]. Xu et al. presented a method for pedestrian detection and tracking using a single night-vision video camera installed on the vehicle [3]. The detection phase was performed by a support vector machine (SVM) with size-normalized pedestrian candidates. In the work of Yasuno et al. [4], the P-tile method was developed to detect the human head first, and then the human torso and legs are included by a local search. Owechko et al. proposed a particle swarm optimization algorithm for human detection in IR imagery [5]. Davis et al. presented a two-stage template-based method with an Adaboosted classifier for pedestrian detection [6]. Dai et al. presented an approach toward pedestrian detection and tracking from infrared imagery using joint shape and appearance cues [7]. A shape cue is first used to eliminate non-pedestrian moving objects and then an appearance cue helps to locate the exact position of pedestrians. Li et al. presented a robust pedestrian detection method in thermal infrared images based on the double-density dual-tree complex wavelet transform (DD-DT CWT) and wavelet entropy [8]. In the work of Wang et al. [9], the GMM background model was first deployed to separate the foreground candidates from the background, then a shape describer was introduced to construct the feature vector for pedestrian candidates, and a SVM classifier was trained to detect the pedestrian. Ko et al. introduced an efficient human detection method in thermal images, using a center-symmetric local binary pattern (CS-LBP) with a luminance saliency map and a random forest (RF) classifier scheme [10].

Great progress has been made as reported in various literature, however, practical pedestrian detection still suffers from a lack of robustness. To address this issue, in this paper, we propose a robust pedestrian detection method by using the shape

* Corresponding author. Tel.: +86 15657104670.
E-mail address: zaixinghe@zju.edu.cn (Z. He).

http://dx.doi.org/10.1016/j.patcog.2014.12.013
0031-3203/© 2014 Elsevier Ltd. All rights reserved.

distribution histogram (SDH) feature with a contour saliency map and modified sparse representation classification (MSRC) in the thermal infrared imagery. Since the brightness is unreliable in the thermal infrared imagery, the proposed method first utilizes the contour saliency map and a segmentation method to detect candidate pedestrian regions. Then distances between random points on the thinned contour map of objects in the candidate regions are applied to acquire the SDH feature, which can precisely describe the pedestrian shape. Finally, a robust MSRC classifier which has high accuracy is used to recognize the true pedestrians using the SDH feature.

The main contribution of the proposed paper is the robust SDH feature for infrared pedestrian detection. Nowadays, various features can be used for representing objects, such as color, texture and shape [11]. Since the infrared images always have lower spatial resolutions and cannot provide as much information as visible ones about objects, shape cues are more reliable and particularly attractive in an infrared imaging detection system [12–14]. In this study, we propose a robust and discriminative shape distribution histogram feature for describing shapes of objects, which was inspired by the work of Osada et al. [15]. The SDH feature uses random distance sampling to produce a continuous probability distribution. It can discriminate different classes of objects correctly despite the object size and small pose changes. Furthermore, it is simple and fast and can be used in real-time detection. In addition, we propose a MSRC classifier which is based on the original SRC classifier [16]. Compared with the original one, MSRC is more robust in recognizing pedestrians and non-pedestrians.

The remainder of the paper is organized as follows. Section 2 presents the candidate regions detection. Section 3 introduces the SDH feature and the MSRC classifier to recognize pedestrians and non-pedestrians. Experimental results for infrared pedestrian detection under various scenarios are shown in Section 4. Finally, we conclude the paper in Section 5 with a summary.

2. Candidate region detection

We begin the process by detecting the candidate regions in images that are likely to contain foreground objects. Most of pedestrian detection methods in thermal infrared images assume that pedestrians are warmer than the background. These methods select regions that are hotter than the background using the threshold [10]. It works in most of the time, especially at night and during the winter. However, these techniques are unable to detect pedestrians which are colder than the background, a situation that can happen in hot summer days. Although using adaptive thresholds may improve the performance [17], different from them, we adopt the contour saliency map to detect the contour boundary of objects, which is unaffected by temperature differences. This part is introduced in Section 2.1. Furthermore, in Section 2.2, we introduce a segmentation method based on the intensity vertical projection to obtain the separate candidate region of people.

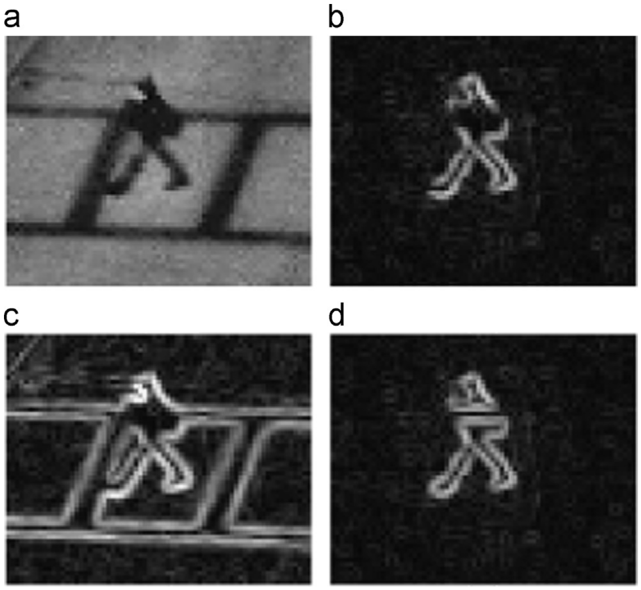

Fig. 1. The contour saliency map. (a) The original image, (b) the contour saliency map, (c) the visualization of $\|\nabla I(i,j)\|$, and (d) the visualization of $\|\nabla(B(i,j)-I(i,j))\|$.

2.1. Contour saliency map

A contour saliency map [18] of a thermal image represents the probability of each pixel belonging to the contour boundary of foreground objects. It only shows the gradients in the input image that are both strong and significantly different from the background. Thus, false lines around the boundary of foreground objects can be removed successfully, which is shown in Fig. 1.

The input and background gradient information are combined to get the contour saliency map of objects. For one input image $I(i,j)$, its Contour Saliency Map is indicated as

$$C(i,j)=\min\left(\|\nabla I(i,j)\|,\|\nabla(B(i,j)-I(i,j))\|\right), \tag{1}$$

where $B(i,j)$ can be obtained as the mean intensity of the training data. Fig. 1(b) is the Contour Saliency Map of Fig. 1(a). Fig. 1(c) is the visualization of $\|\nabla I(i,j)\|$, and Fig. 1(b) shows that of $\|\nabla(B(i,j)-I(i,j))\|$.

2.2. Pedestrian segmentation

After the contour saliency map has been obtained from the infrared image, we need to threshold the contour saliency map into a binary image to select the most salient contours. Then a mathematical morphology operator is employed to reduce the noise, and the foreground connected region is extracted and fitted with a rectangle bounding box. In the case of single pedestrian, this rectangle bounding box can be defined as the candidate region of the human body.

Infrared images generally have noise and low brightness pixels. This may limit the effect of the segmentation of the human. Due to the pixels on the human body which have low contrast, a single pedestrian may sometimes be divided by several bounding boxes.

To solve this issue, we first fuse bounding boxes whose weight center points have a close Euclidean distance. The distance parameter is related to the size of the pedestrian in the image. It should be selected appropriately. If the parameter is too small, the divided people may not be fused to a single one, otherwise, close people may be fused incorrectly. Since we will use a segmentation method in the next step, a larger distance parameter would increase the detection time, but has no obvious influence on the detection accuracy. According to our experimental experiences, we suggest the distance parameter is not smaller than the one-third of the diagonal length in the pedestrian template. In our experiments, we used half of the diagonal length.

The pedestrian template size can be decided based on the priori knowledge. If pedestrians have very small size differences in the whole image, a single template can be used. If their sizes are different due to the linear perspective, we choose different templates for different regions. Generally, pedestrians close to the camera are relatively large and at the bottom of the image, and pedestrians far from the camera are usually small and at the top. In our experiment, two templates are used for the upper and bottom half regions of the image. The standard template size in each region is determined from the manually detected pedestrians in the training images. 16 × 30 and 18 × 36 are used as the template size in the experiment.

Then a segmentation method is utilized to separate close people that are fused incorrectly and then we can differentiate overlapped people. The human head in the infrared image always has an obvious characteristic. Thus, we can estimate the number of pedestrians in the bounding box by considering their vertical projections [9]. Fig. 2 shows that the peak in the vertical projection curve usually represents the position of the human head. The peak value can be obtained by comparing each element of projection data to its neighboring values. If an element is larger than both of its neighbors, it is chosen as the local peak. In addition, to reduce the error, minimum peak height is set as a half of the template height, and the minimum separation between peaks is selected as a quarter of the template width. When there are multiple local peaks within a certain distance (the defined minimum separation between peaks), the highest one is chosen as the final peak.

We mark each peak as the center line of the candidate region. The transverse distance of the pedestrian is difficult to decide because of the overlap. So based on the center line, the left and right borders of the single pedestrian are obtained using a standard template width for simplicity. In the case that the camera is not so close to the object, size differences of pedestrians in the image are small, so that the error caused by the standard template is also small. Template width in each region can be valid for pedestrians of different sizes. The longitudinal distance of the pedestrian can be decided more exactly and reliably. The top and bottom borders of the single pedestrian are calculated by considering the intersection of the center line and the human body.

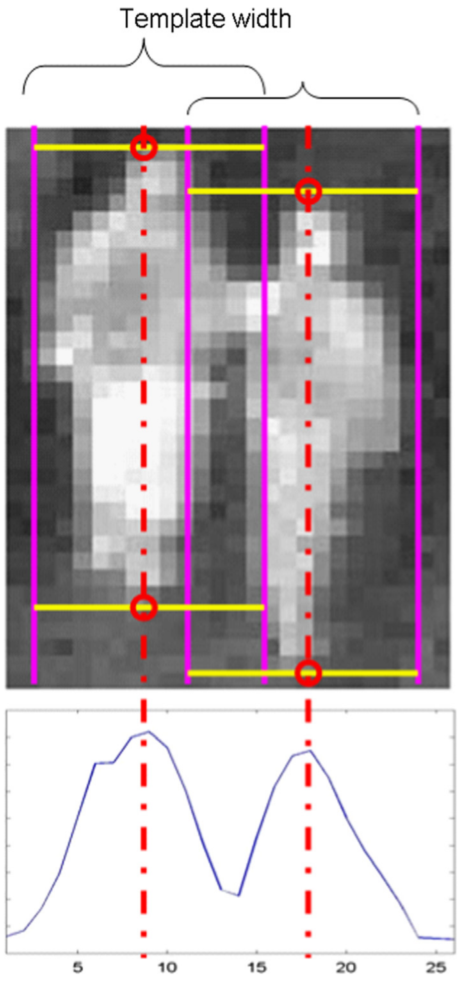

Fig. 2. The peak in the vertical projection curve usually represents the position of the human head.

It is obvious that the candidate selection method results in bounding boxes with different width/height ratios. Since the shapes and sizes of the real people are not the same, the changeable width/height ratio is more reasonable and has smaller influence on detection performance than the fixed width/height ratios. Some examples of the candidate region segmentation results are shown in Fig. 3. The yellow rectangles are the original segmentations. The blue rectangles show the results after fusion, and the red rectangles are the final results.

3. Pedestrian classification

After candidate regions have been acquired, these regions will be classified and the real pedestrians will be recognized during further verification. A discriminative pedestrian feature and an effective classification technique are needed in this stage. In this section, we first introduce the SDH feature that describes the shape of a human body using the shape distance probability distributions, then a robust MSRC classifier is trained to differentiate pedestrians and non-pedestrians.

3.1. The SDH feature

3.1.1. Description of SDH

The feature extraction is the core part of a pedestrian classification. The shape-based feature is more useful for pedestrian detection in infared images since it usually provides more reliable information in the general case. In this work, we propose a robust and efficient shape representation feature motivated by the work of Osada et al. [15] which shows excellent performance in graphic matching. The feature is simple and can effectively discriminate objects with different shapes. In addition, it has several desirable properties for the applications in this work. First, it yields invariance under motion and scale by using the distance histogram distribution. Second, it is robust to noise and blur since random sampling of it ensures that shape distributions are insensitive to small perturbations. Third, it is fast and efficient which can be used in real-time detection.

In the framework, the contour saliency map is first thinned to one-pixel thick contours. The contour saliency map has the composite characteristic, so that the maximum in it always co-occurs with the maxima in the input gradients. Therefore, we use the non-maximum suppression result of the input gradients as the thinning mask. The thinned contour saliency map is denoted as

$$S(i,j) = C(i,j) \times f(\| \nabla I(i,j) \|), \tag{2}$$

in which, $f(\cdot)$ is the non-maximum suppression function.

After thinning, we select the salient contour segments from $S(i,j)$ which can represent the boundary of the human body well as

$$B(i,j) = \begin{cases} 1 & \text{if } S(i,j) \geq \omega, \\ 0 & \text{otherwise}, \end{cases} \tag{3}$$

where ω is the threshold that was determined by using the Otsu method [19]. Fig. 4 shows the process to get the boundary of the human body. Fig. 4(b) is the contour saliency map of Fig. 4(a). Fig. 4(c) is the thinned contour saliency map, and Fig. 4(d) represents the boundary of the human body.

Based on it, a geometric shape function D_k is distributed by measuring the Euclidean distance between two random points from the pixels of the candidate regions. Let the kth point pair be (i_{k1}, j_{k1}) and (i_{k2}, j_{k2}), where $B(i_{k1}, j_{k1}) = B(i_{k2}, j_{k2}) = 1$. D_k is denoted

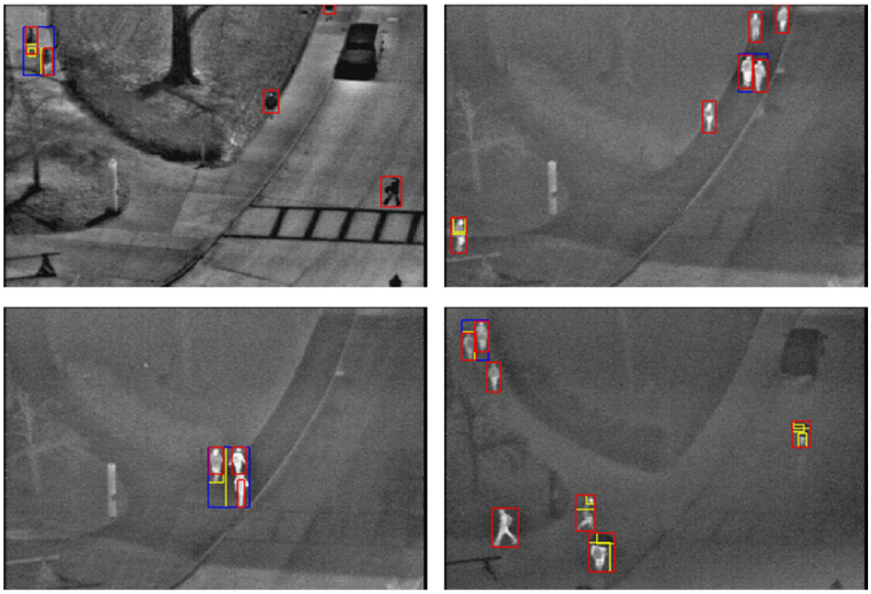

Fig. 3. Examples of the candidate region detection results. The yellow rectangle shows the original segmentation. Single pedestrian is possibly divided by several bounding boxes due to the separation of foreground regions. The blue rectangle shows the results after fusion, and the red rectangle shows the final results. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

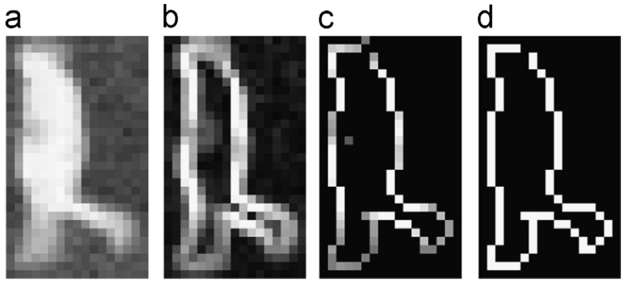

Fig. 4. The process to get the boundary of human body. (a) The test subject; (b) the contour saliency map; (c) the thinned contour saliency map; and (d) the boundary of the human body.

as

$$D_k = \sqrt{(i_{k1}-i_{k2})^2+(j_{k1}-j_{k2})^2}. \tag{4}$$

Then, K samples evaluated from the shape distribution D_k and a histogram is constructed by counting how many samples fall into each of the m bins. From the histogram, we reconstruct a piecewise linear function with equally spaced vertices, which forms the representation for the SDH feature. We compute the SDH feature for each candidate regions and store it as a sequence of integers. For the fair comparison, all distance distributions are normalized to a standard value. Thus, the SDH feature $H \in \mathbb{R}^m$ is indicated as

$$H(i) = \frac{1}{K}\sum_{k=1}^{K} F_i(D_k), \tag{5}$$

where

$$F_i(D_k) = \begin{cases} 1 & if\ \frac{i-1}{m} < \frac{D_k}{\max\limits_j(D_j)} \le \frac{i}{m}, \\ 0 & otherwise. \end{cases} \tag{6}$$

Here, $i \in \{1,\ldots,m\}$ and $j \in \{1,\ldots,K\}$.

3.1.2. SDH properties

The SDH is easy to compute and is invariant to scale changes and small poses. The SDH feature has low computational complexity. The calculation of the SDH feature contains computing distances of K point pairs and quantizing the K distances to the bins of the histogram. The distance calculation and the quantization require a small quantity of calculation and the complexity of these items is constant. Thus, the computational complexity of the SDH feature is $O(K)$.

SDH is insensitive to small pose changes. For the 3D object, the point pair distance is constant no matter what pose the object has. The 2D image is the projection of the 3D object from an angle of view. In other words, the 2D objects with different poses are the projections of the 3D object in different views. The small changes of angle of the view cause the small changes of the point pair distance and thereby changes of the distant histogram are not obvious. Thus, the SDH feature is resistant to small pose changes. For example, the distance for a randomly selected point pair (i_1,j_1) and (i_2,j_2) is indicated as $D=\sqrt{(i_1-i_2)^2+(j_1-j_2)^2}$ in the projected 2D image. Assuming that the rotation angle of the pedestrian in the x coordinate is α, then the pair distance can be denoted as $D^* = \sqrt{(i_1-i_2)^2 \cdot \cos^2\alpha+(j_1-j_2)^2}$. It can be seen that when the α is small, the change of the distance is small. Similar results can be obtained for the other two rotation angles.

SDH is also robust to scaling. Given a scaling factor μ, each pair distance is changed from D to $\mu \cdot D$. Then the threshold $D_k/\max_j(D_j)$ in Eq. (6) is changed into $\mu \cdot D_k/\max_j(\mu \cdot D_j) = \mu \cdot D_k/\mu \cdot \max_j(D_j) = D_k/\max_j(D_j)$, which is the same with that before scaling. Thus, the distance histogram remains the same.

We test the similarity of the SDH feature for the pedestrian regions in Fig. 6. Fig. 6(a) shows sample pedestrian regions, in which pedestrians have different poses and scales, and the intensities of the pedestrians are obviously different because of the temperature changes. Fig. 6(c) demonstrates the similarity among the SDH feature curves for pedestrian regions in Fig. 6(a). For comparison, we also draw the brightness-histogram-curves of Fig. 6(a) in Fig. 7, which is introduced in [20]. It is shown that the proposed SDH feature is much more robust than the brightness histogram in representing the similarity of pedestrians.

We also test the classification ability of the SDH feature. Fig. 6(b) shows some sample images of non-pedestrian regions, and Fig. 6(d) shows the SDH feature curves in the regions of Fig. 6(b). The comparison between Fig. 6(c) and (d) reveals that the SDH feature can differentiate the pedestrians and non-pedestrians very well.

3.1.3. *Parameter discussions*

Parameters K and m in the SDH feature are discussed. K is the number of point pair samples for the shape distribution. The more the samples are picked, the better the result describes the real situation. However, large K will increase the calculation time. Thus, the appropriate value of K should be picked to make the balance between the accuracy and efficiency. K is picked from 50 to 10,000 to test the stability of the SDH feature in Fig. 5(a). It can be seen that the larger the value of K, the smoother are the feature curves. On the other hand, when K is set large enough ($K \geq 1000$), the feature curve can keep a stable shape. In our experiment, K is set at 1000.

Parameter m describes the number of histogram bins. Theoretically, the number of histogram bins indicates the dimension of the feature vector. The smaller the dimension, the less complicated are the computer calculations. But at the same time, the discrimination of the feature may decrease. We use the relative similarity S to measure the feature discrimination as follows:

$$S=\frac{1}{C(C-1)}\sum_{u=1}^{C}\sum_{v=1,v\neq u}^{C}H_u^{\mathrm{T}}H_v-\frac{1}{C^2}\sum_{u=1}^{C}\sum_{w=1}^{C}H_u^{\mathrm{T}}H_w, \tag{7}$$

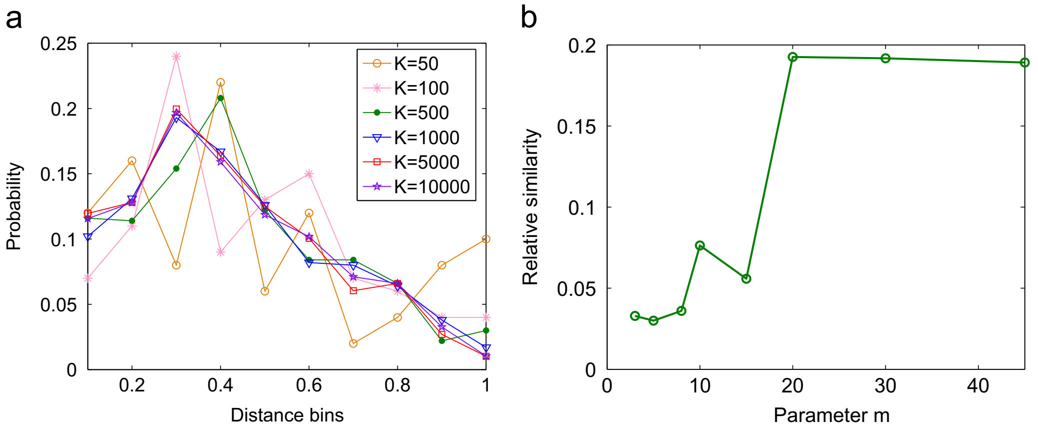

Fig. 5. Parameter discussion. (a) The parameter K; (b) the parameter m.

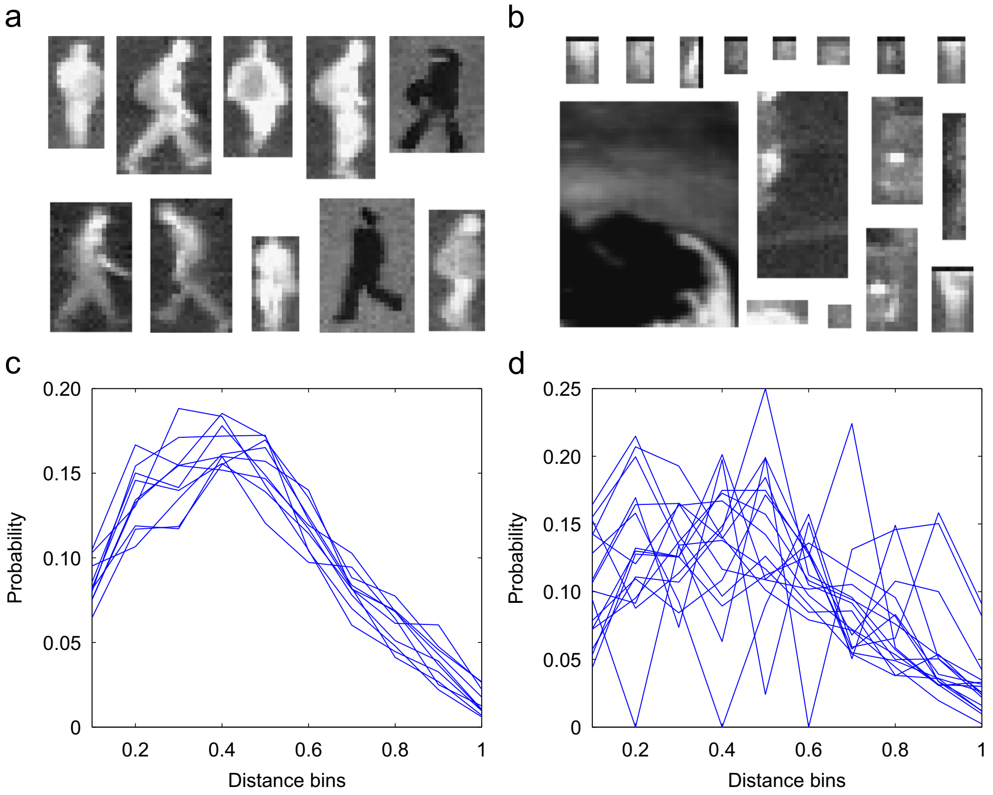

Fig. 6. Properties of the SDH feature. (a) The sample regions of pedestrians, in which pedestrians have different poses and scales, and the intensities of pedestrians are obvious different because of the temperature changes. (b) The sample regions of non-pedestrian. (c) The similarity of the SDH feature curve in the pedestrian region of (a). (d) The SDH feature curves in the non-pedestrian region of (b), which represents the classification ability of the SDH feature.

where H_u and H_v indicate the SDH features for people, H_w is for the non-people, and C is the number of test samples of people (for simplicity, that of non-people is also C). Thus, $\frac{1}{C(C-1)}\sum_{u=1}^{C}\sum_{v=1,v\neq u}^{C} H_u^{\mathrm{T}} H_v$ shows the average intraclass similarity (similarity between people), while $\frac{1}{C^2}\sum_{u=1}^{C}\sum_{w=1}^{C} H_u^{\mathrm{T}} H_w$ shows the average interclass similarity (similarity between people and non-people). From Eq. (7), it can be seen that the larger the relative similarity becomes, the easier the classifier identifies. It is better to choose a suitable dimension value to ensure that the relative similarity decreases notably when reducing the dimension, and no significant changes of relative similarity appear when increasing the dimension. To pick a good m, we select m = 3, 5, 8, 10, 15, 20, 30, 45 and C=50, and draw the relative similarity curve in Fig. 5(b). It is shown that when parameter m is smaller than 20, the relative similarity changes rapidly. However, the relative similarity remains constant when m is more than 20. Thus, we choose m=20 in the experiment.

3.2. MSRC classifier

SRC is a pattern classification method, which implements sparse representation of data by using the methods for sparse signal reconstruction in compressed sensing and classifies data in terms of reconstruction errors [16]. SRC was used for robust face recognition to cope with noise corruption, occlusion, outlier detection and etc. It showed excellent classification performance compared with other classifiers, such as the nearest neighbor (NN) [21], nearest subspace (NS) [22,23], and linear SVM [24] in the face databases. However, the original SRC algorithm is not suitable for the classification problem in this work. Thus, we propose a MSRC classifier.

Given a training data set $\mathbf{a}_i \in \mathbb{R}^m$ $(i \in \{1,2,\ldots,n\})$, where each sample $\mathbf{a}_i$ relates a class label l_i $(l_i \in \{1,2,\ldots,c\})$. Let matrix $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n] \in \mathbb{R}^{m\times n}$. Given an error tolerance $\epsilon > 0$ and a test sample $\mathbf{y} \in \mathbb{R}^m$, the SRC algorithm can be summarized as below:

(1) Solve the sparse representation problem via ℓ_1-norm minimization:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1, \quad \text{s.t.} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \le \epsilon \tag{8}$$

(2) Compute the residual $r_j(\mathbf{y}) = \|\mathbf{y} - \mathbf{A}\delta_j(\hat{\mathbf{x}})\|_2$, for $j = 1, \ldots, c$, where $\delta_j : \mathbb{R}^n \to \mathbb{R}^n$ is the characteristic function that selects the coefficients associated with the jth class and sets the others be 0;
(3) Identify $\mathbf{y}$: $I(\mathbf{y}) = \arg\min_i r_i(\mathbf{y})$.

In this work, the classification problem can be viewed as a two-class classification problem. To use the original SRC algorithm, we can choose some pedestrian and non-pedestrian samples to construct the matrix $\mathbf{A}$. Then the following processes are the same as those in face recognition. However, SRC is not very suitable for the problem in this work since it is different from the face recognition problem. In face recognition, the samples can be viewed as nearly equally distributed points in the face subspace. The representational abilities of the samples of each subject are nearly the same. In other words, the samples have nearly no representational bias. However, in the pedestrian classification, the representational abilities of the positive and negative samples are not equal. The positive samples span the subspace of pedestrians, while the negative samples span a wider subspace because of the variety of negative samples. Such unequally distributed samples have a strong representational bias. Therefore, the original SRC algorithm is a bias to classify the test pattern into non-pedestrian. It needs to be modified to fit the target pedestrian recognition problem.

The base of the MSRC is the same as the original SRC algorithm: a test signal can be sparsely represented by the samples belonging to the same pattern. In the proposed algorithm, we do not use any negative samples, which means that $\mathbf{A}$ consists of only positive samples. Furthermore, we find a sparse representation of $\mathbf{y}$ over $\mathbf{A}$. If $\mathbf{y}$ is a pedestrian, the angle between $\mathbf{y}$ and the columns in $\mathbf{A}$ should be small. Therefore, the amplitudes of the solved coefficients $\mathbf{x}$ should be relatively small. Otherwise, those of the coefficients corresponding to a non-pedestrian should be large. Furthermore, as described in [16], another property of sparse representation is that the solved coefficients $\mathbf{x}$ of a pedestrian should be concentrated and those of a non-pedestrian should not be. Fig. 8 shows the examples with 100 positive training samples. Fig. 8(a) shows the coefficient curve for a pedestrian. It can be seen that amplitudes of coefficients are very small (smaller than 0.5). The coefficient with the largest amplitude value represents the most similar training sample. In the case with non-pedestrian which is shown in Fig. 8(b), the amplitudes of coefficients are much larger. Furthermore, the coefficients of the pedestrians in Fig. 8(a) are much more concentrated than those of the non-pedestrians in Fig. 8(b).

The ℓ_1-norm $\|\mathbf{x}\|_1$ can be used to measure the concentration and amplitudes of the coefficients. Firstly, a small ℓ_1-norm stands for concentrated coefficients. With a fixed ℓ_2-norm of 1, the more concentrated the coefficients are, the smaller the ℓ_1-norm will be. Extremely, when the coefficients are the most concentrated (one of the elements is 1 and the others are 0), the ℓ_1-norm has the minimum value of 1. Thus, using the ℓ_1-norm to measure the concentration of the coefficients is an effective approach. Secondly,

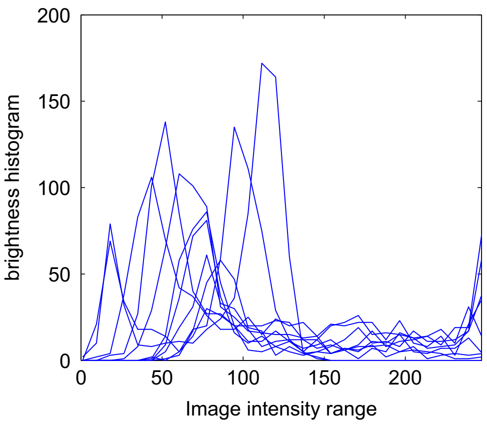

Fig. 7. The brightness-histogram-curves of Fig. 6(a).

Table 1
The quantitative performance for the candidate region detection stage. (#TP: True Positive; #FP: False Positive).

Sequence	# Frame	# Total people	#FP	#TP	PPV (%)	Sensitivity (%)
1	28	80	9	79	88.75	98.75
2	25	85	6	85	92.94	100.00
3	20	88	13	88	85.23	100.00
4	15	95	15	94	84.21	98.95
5	20	95	6	95	93.68	100.00
6	15	74	7	74	90.54	100.00
7	19	88	11	85	87.50	96.59
8	21	87	16	81	81.61	93.10
9	70	92	5	92	94.57	100.00
10	21	80	16	80	80.00	100.00
Total	254	864	104	853	87.96	98.73

a small ℓ_1-norm also generally stands for small amplitudes of the elements. The smaller the amplitudes of the coefficients are, the smaller the ℓ_1-norm will be. Therefore, the pedestrians and non-pedestrians can be classified by a ℓ_1-norm threshold after sparse representation. Actually, the ℓ_1-norms of pedestrians and non-pedestrians are highly discriminative. For example, the ℓ_1-norm of the pedestrians in Fig. 8(a) is only 1.40, while that of the non-pedestrians in Fig. 8(b) is 72.78.

Let matrix $\mathbf{A}=[\mathbf{a}_1,\mathbf{a}_2,\ldots,\mathbf{a}_n]\in\mathbb{R}^{m\times n}$, where $\mathbf{a}_i$ is a positive sample of pedestrian. Given a test sample $\mathbf{y}\in\mathbb{R}^m$, the proposed MSRC algorithm can be summarized as follows:

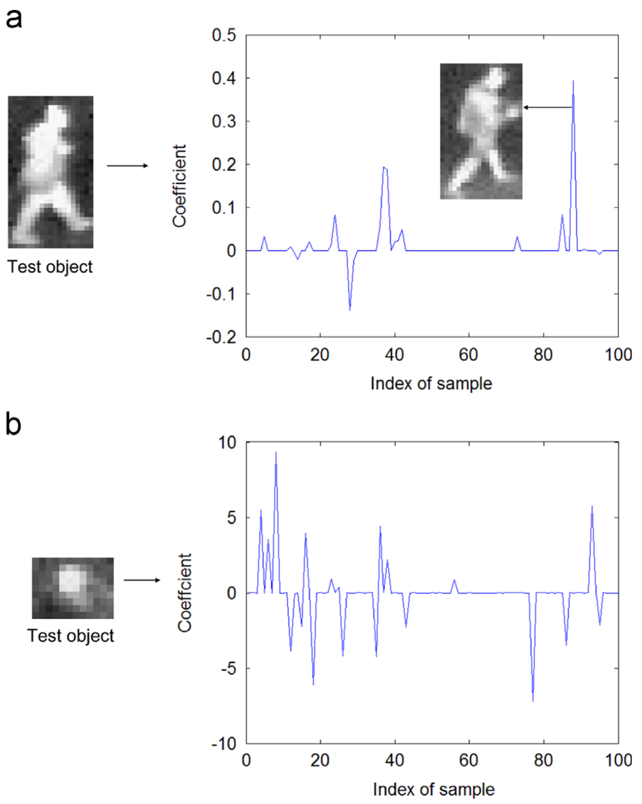

Fig. 8. Sparse coefficient curves for example test objects. (a) The case with a pedestrian. The coefficients **x** is concentrated and the amplitudes are small. The coefficient with the largest amplitude value represents the most similar training sample. (b) The case with a non-pedestrian. The amplitudes of coefficients are large.

(1) Solve the ℓ_1-norm minimization of (8);
(2) Compute the ℓ_1-norm $\|\hat{\mathbf{x}}\|_1$, the solution of Step 1);
(3) Identify $\mathbf{y}$: $I(\mathbf{y})=1$ if $\|\hat{\mathbf{x}}\|_1\le\lambda$, otherwise $I(\mathbf{y})=0$.

The parameter λ can be chosen through a training process as follows. For each positive sample $\mathbf{a}_i$, find its sparse representation over the other samples as follows:

$$\hat{\mathbf{x}}_i=\arg\min_{\mathbf{x}}\|\mathbf{x}\|_1,\quad \text{s.t.}\|\overline{\mathbf{A}}_i\mathbf{x}-\mathbf{a}_i\|_2\le\epsilon, \tag{9}$$

where $\overline{\mathbf{A}}_i$ consists of the columns of $\mathbf{A}$ except $\mathbf{a}_i$. Since the ℓ_1-norms of the pedestrians and non-pedestrian are highly discriminative, a simple way for determining λ can be denoted as

$$\lambda=\alpha\cdot\max_i\|\hat{\mathbf{x}}_i\|_1, \tag{10}$$

where α is a relaxation factor. Based on the numerical studies, good performance can be achieved when α is in the range of [1, 2]. In our experiment, α is set as 1.5.

4. Experimental results

In the experiment, we test the performance of the proposed method in a OTCBVS Benchmark Dataset Collection—OSU thermal pedestrian database [25]. There are 10 test sequences in the OSU thermal database, which covers a variety of environmental conditions such as rainy, cloudy and sunny days. In the experiment, 1000 randomly selected point pairs are picked to obtain the SDH feature. The feature dimension of the SDH feature is set as 20 for the classification, and λ is picked as 5 in the MSRC. The proposed algorithm consists of three components—segmentation, feature extraction and classification. We use the C++ language to implement the first two steps and use the MATLAB to implement the third part. Programming in the mixed mode with MATLAB and C++, our algorithm runs about 31 frames/s on an Intel 3.10 GHz CPU.

4.1. Results of the candidate region detection stage

In the proposed method, two different stages are composed to solve the detection problem and the classification problem independently. The performance of both of the stages are evaluated in this work.

We firstly test the quantitative performance for the candidate region detection stage in Table 1. Sensitivity and PPV are

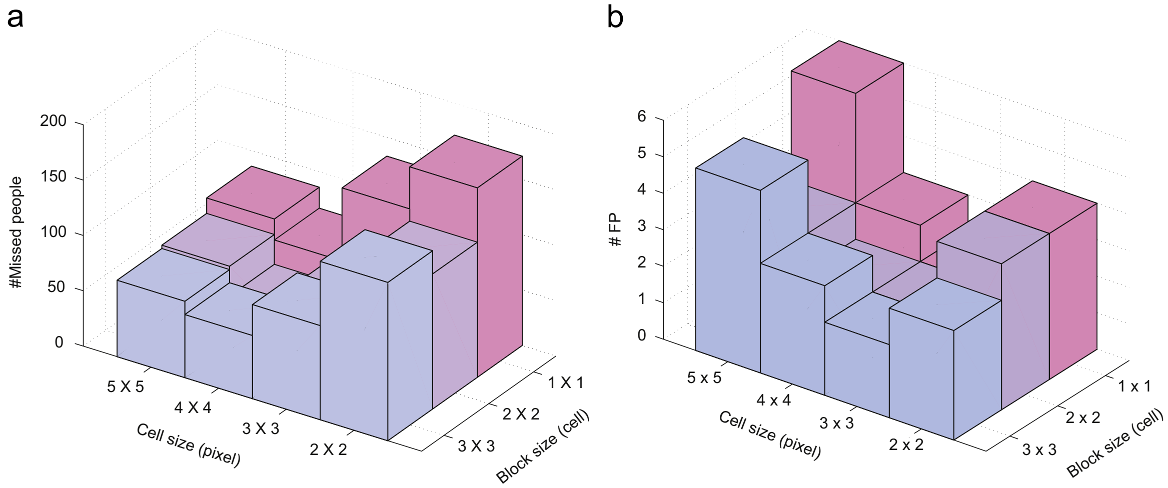

Fig. 9. The performance of HOG+SVM in different HOG cell sizes and block sizes. (a) The #Missed people (# Total people−#TP). (b) shows the #FP (False Positive number).

Table 2
Comparison results for the classification stage with different features. Note that the number of the people is the number of the detected people after the candidate detection stage (#TP: True Positive; #FP: False Positive).

Sequence	# Frame	# People	INERTIA+MSRC				LBP+MSRC				HOG+MSRC				Proposed			
			#FP	#TP	PPV (%)	Sensitivity (%)	#FP	#TP	PPV (%)	Sensitivity (%)	#FP	#TP	PPV (%)	Sensitivity (%)	#FP	#TP	PPV (%)	Sensitivity (%)
1	28	79	0	74	100.00	93.67	0	77	100.00	97.47	0	77	100.00	97.47	0	78	100.00	98.73
2	25	85	1	81	98.82	95.29	1	81	98.82	95.29	0	83	100.00	97.65	0	85	100.00	100.00
3	20	88	1	88	98.86	100.00	1	73	98.86	82.95	0	80	100.00	90.91	0	87	100.00	98.86
4	15	94	2	71	97.87	75.53	3	72	96.81	76.60	2	90	97.87	95.74	0	94	100.00	100.00
5	20	95	1	79	98.95	83.16	0	81	100.00	85.26	0	92	100.00	96.84	1	95	98.95	100.00
6	15	74	0	72	100	97.30	0	69	100.00	93.24	0	73	100.00	98.65	0	74	100.00	100.00
7	19	85	2	60	97.65	70.59	1	77	98.82	90.59	1	85	98.82	100.00	0	81	100.00	95.29
8	21	81	11	48	86.42	59.26	2	67	97.53	82.72	1	76	98.77	93.83	0	81	100.00	100.00
9	70	92	0	92	100.00	100.00	0	90	100.00	97.83	0	91	100.00	98.91	0	92	100.00	100.00
10	21	80	2	78	97.50	97.50	1	70	98.75	87.50	2	76	97.50	95.00	0	80	100.00	100.00
Total	254	853	20	743	97.66	87.10	9	757	98.94	88.75	6	823	99.30	96.48	1	847	99.88	99.30

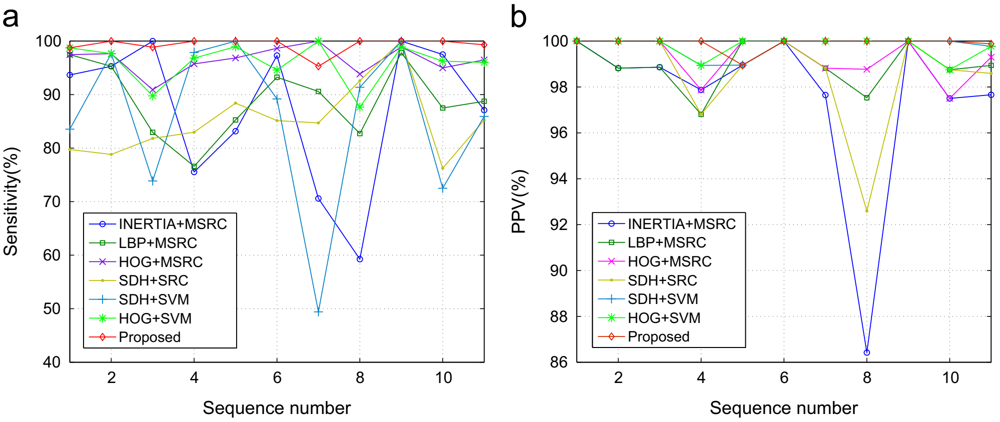

Fig. 10. Comparisons of PPV and Sensitivity rates for different algorithms in the classification stage (sequence #11 is for the total sequence). (a) The Sensitivity rate. (b) The PPV rate.

Table 3
Comparison results for the classification stage with different classifiers. Note that the number of the people is the number of the detected people after the candidate detection stage (#TP: True Positive; #FP: False Positive).

Sequence	# Frame	# People	SDH+SRC				SDH+SVM				HOG+SVM				Proposed			
			#FP	#TP	PPV (%)	Sensitivity (%)	#FP	#TP	PPV (%)	Sensitivity (%)	#FP	#TP	PPV (%)	Sensitivity (%)	#FP	#TP	PPV (%)	Sensitivity (%)
1	28	79	0	63	100.00	79.75	0	66	100.00	83.54	0	78	100.00	98.73	0	78	100.00	98.73
2	25	85	0	67	100.00	78.82	0	83	100.00	97.65	0	83	100.00	97.65	0	85	100.00	100.00
3	20	88	0	72	100.00	81.82	0	65	100.00	73.87	0	79	100.00	89.77	0	87	100.00	98.86
4	15	94	3	78	96.81	82.98	1	92	98.94	97.87	1	91	98.94	96.81	0	94	100.00	100.00
5	20	95	1	84	98.95	88.42	1	95	98.95	100.00	0	94	100.00	98.95	1	95	98.95	100.00
6	15	74	0	63	100.00	85.14	0	66	100.00	89.19	0	70	100.00	94.59	0	74	100.00	100.00
7	19	85	1	72	98.82	84.71	0	42	100.00	49.41	0	85	100.00	100.00	0	81	100.00	95.29
8	21	81	6	75	92.59	92.59	0	74	100.00	91.36	0	71	100.00	87.65	0	81	100.00	100.00
9	70	92	0	92	100.00	100.00	0	92	100.00	100.00	0	91	100.00	98.91	0	92	100.00	100.00
10	21	80	1	61	98.75	76.25	0	58	100.00	72.50	1	77	98.75	96.25	0	80	100.00	100.00
Total	254	853	12	727	98.59	85.23	2	733	99.77	85.93	2	819	99.77	96.01	1	847	99.88	99.30

used to estimate the performance of the proposed method quantitatively [7],

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{Pedestrians in total}}, \tag{11}$$

$$\text{PPV} = 1 - \frac{\text{False Positive}}{\text{Pedestrians in total}}. \tag{12}$$

The Sensitivity reports the probability of people that are correctly identified, where a high Sensitivity value corresponds to a high detection rate of people. The PPV describes the fraction of detections that actually are people, where a high PPV corresponds to a low number of false positives. It is shown that the proposed candidate region detection method achieved 87.96% PPV rate and 98.73% Sensitivity rate. In the detection stage, totally 11 pedestrians are not detected and 104 non-pedestrians are falsely detected as pedestrians.

4.2. Results of the classification stage

Then the performance of the proposed method in the classification stage is evaluated. In the database, the first three frames in each sequence are used for training. The proposed MSRC classifier is trained by using only the positive samples, so we picked totally

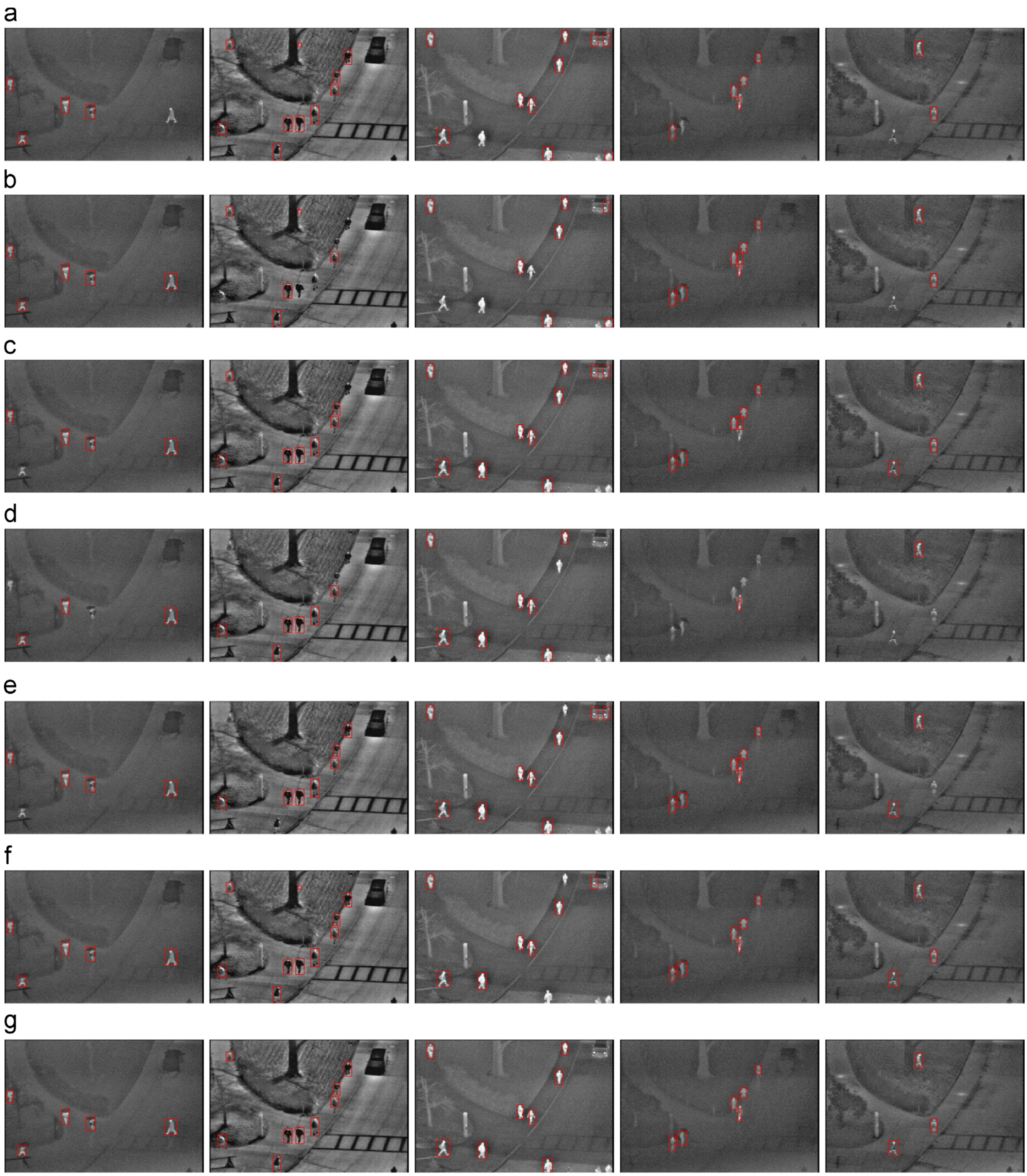

Fig. 11. Some examples of the comparison results. (a) The results of INERTIA+MSRC algorithm; (b) the results of LBP+MSRC algorithm; (c) the results of HOG+MSRC algorithm; (d) the results of SDH+SVM algorithm; (e) the results of SDH+SRC algorithm; (f) the results of HOG+SVM algorithm;(g) the results of the proposed method.

50 positive samples for training. For fair comparison, the comparative classifiers used the same 50 positive samples and other 50 negative samples in the training stage. The rest of the frames are used for testing.

To test the robustness of the proposed method, we change our feature and classifier by some competitive ones respectively. We first compare the proposed method with three different features—intensity distribution based inertia (INERTIA)[20], local binary patterns (LBP) [26], and histogram of oriented gradients (HOG) [27] features, by using the MSRC classifier. INERTIA is a well-known feature which is based on the inertial similarity among pedestrian regions in thermal pedestrian detection. LBP and HOG are two famous feature descriptors used in computer vision and image processing for the purpose of object detection. LBP is a simple but very efficient texture operator. In the comparative test, an (8, 1) neighborhood is used in LBP with uniform patterns,

Table 4
Comparison results of different algorithms using 10 positive and 10 negative training samples.

Sequence	INERTIA+MSRC		LBP+MSRC		HOG+MSRC		SDH+SRC		SDH+SVM		HOG+SVM		Proposed	
	PPV (%)	Sensitivity (%)	PPV (%)	Sensitivity (%)	PPV (%)	Sensitivity (%)	PPV (%)	Sensitivity (%)	PPV (%)	Sensitivity (%)	PPV (%)	Sensitivity (%)	PPV (%)	Sensitivity (%)
1	98.75	82.50	98.75	66.25	100.00	73.75	98.75	75.00	98.75	82.50	98.75	76.25	98.75	88.75
2	96.47	64.71	98.82	87.06	98.82	75.29	98.82	76.47	98.82	81.18	98.82	76.47	100.00	83.53
3	98.86	76.14	95.45	51.14	98.86	73.86	96.59	76.14	96.59	75.00	98.86	69.32	100.00	80.68
4	93.68	67.37	96.84	50.53	91.58	76.84	94.74	78.95	94.74	69.47	93.68	73.68	94.74	82.11
5	95.79	66.32	95.79	60.00	97.89	76.84	95.79	67.37	97.89	71.58	97.89	74.74	97.89	85.26
6	95.95	79.73	95.95	100.00	97.30	82.43	95.95	82.43	98.65	74.32	95.95	94.59	98.65	93.24
7	89.77	61.36	97.73	85.23	96.59	72.73	93.18	68.18	97.73	47.73	96.59	69.32	97.73	72.73
8	88.51	37.93	91.95	43.68	98.85	75.86	89.66	73.56	91.95	77.01	98.85	74.71	98.85	77.01
9	100.00	100.00	100.00	91.30	100.00	97.83	100.00	97.83	98.91	97.83	100.00	96.74	100.00	100.00
10	85.00	75.00	98.75	80.00	97.50	78.75	83.75	73.75	100.00	72.50	96.25	77.50	100.00	90.00
Total	94.33	70.95	96.99	70.83	97.69	78.47	94.79	76.97	97.34	74.88	97.57	78.13	98.61	85.19

yielding 53 different histogram labels. HOG descriptor focuses mainly on silhouette contours. The HOG parameters were adopted after a set of experiments performed with the SVM classifier. Fig. 9 plots both of the #Missed people (#Total people-#TP) and the #FP (False Positive number) in different HOG cell sizes and block sizes. The HOG window size is fixed as the bounding box size, the block overlap is set at half of the block size and the histogram channel is set as 9. It is shown that 2×2 block size with 4×4 cell size performs the best, achieving 45 of #Missed people and 2 of #FP in our experiment.

Table 2 summarizes the Sensitivity and PPV in 10 sequences of the OSU thermal pedestrian database with different features. We only evaluate the classification performance, so all the results are based on the first stage. The number of the people in Table 2 is the number of the detected people after the detection stage. To ensure the fair comparisons and repeatability, a standard criteria–Pascal Criteria is used [28]. The ground truth bounding box is marked manually and a pedestrian with more than 40% occlusion is considered to be a non-pedestrian. Then a overlap probability ζ between the predicted bounding box B_p and the ground truth bounding box B_g is indicated as

$$\zeta=\frac{\mathrm{area}(B_p\cap B_g)}{\mathrm{area}(B_p\cup B_g)}, \tag{13}$$

where $B_p\cap B_g$ denotes the intersection of the predicted and ground truth bounding box, and $B_p\cup B_g$ denotes their union. When ζ exceeds 60%, the recognition is considered a correct recognition.

The INERTIA feature achieves a 97.66% PPV rate and 87.10% Sensitivity rate in the total sequences. The detection result of the LBP feature is 98.94% in PPV and 88.75% in Sensitivity, and the HOG feature shows a 99.30% PPV rate and 96.48% Sensitivity rate. Comparing with these methods, we get a Sensitivity rate of 99.30% (6 pedestrians are missed), and a PPV rate of 99.88% (1 falsely detected pedestrian) using the proposed SDH feature, which is better than the INERTIA, LBP, and HOG features.

We then compare the proposed method with two famous classifiers—the standard SRC and SVM, using the SDH feature. Since the HOG feature is usually used with the SVM classifier [27], this well-established combination is also compared with the proposed method. Table 3 summarizes the classification results. The MSRC classier achieves better performance results (99.88% in PPV and 99.30% in Sensitivity) than both of the standard SRC classier (98.59% in PPV and 85.23% in Sensitivity) and the SVM classifier with the SDH feature (99.77% in PPV and 85.93% in Sensitivity). It is also shown that the HOG and SVM combination can achieve high PPV (99.77%) and Sensitivity (96.01%), but it is not as good as the proposed algorithm. Fig. 10 shows more straightforward comparisons of PPV and Sensitivity rates in the classification stage.

Furthermore, some sample results after the two stages are shown in Fig. 11. Row (a) is the result of the INERTIA+MSRC algorithm, row (b) is that of the LBP+MSRC algorithm, row (c) is that of the HOG+MSRC algorithm, row (d) shows the results of the SDH+SVM algorithm, row (e) shows the results of the SDH+SRC algorithm, row (f) shows the results of the HOG+SVM algorithm, and row (g) shows the result of the proposed method. It can be seen that the proposed method showed good performance in detecting pedestrians, and the comparison algorithms could not either remove the non-pedestrian subjects or detect the pedestrian subjects correctly.

In addition, the proposed method is compared with the state-of-art work in thermal pedestrian detection [6]. In [6], Davis et al. proposed a famous two-stage template-based method. Furthermore, the detection result for the same challenging database was demonstrated in the paper. For the Davis's method, PPV of the 10 sequence is 99.39% and the Sensitivity is 94.51%. Our approach works better on both of Sensitivity and PPV.

4.3. Discussion of training numbers

In this subsection, we discuss the performance of the algorithms with different numbers of training samples. We picked 10, 20, 50 and 100 positive samples, and 10, 20, 50 and 100 negative samples in the first three frames independently. The final results including the candidate region detection and the classification stages are shown in Tables 4–7. Figs. 12–15 corresponding to the tables are also included for easy understanding. In addition, Fig. 16 illustrates Sensitivity and PPV performances of the seven different algorithms with the changing sample numbers. The proposed method showed better performance by comparing to the other methods.

For a classifier, the training process is to learn the pattern with some samples. The recognition performance is affected by whether the samples are sufficient or not. Before the samples are sufficient, the recognition rate can raise greatly with the increase of the sample number. When the sample size is large enough, there will be no obvious improvement in the recognition rate. In Fig. 16, it can be seen that the classifiers achieve relatively very high performances with 50 training samples. When the training number is smaller than 50, the recognition performances of algorithms become better with the increase of training samples. However, when the training number is larger than 50, the PPV and Sensitivity changes are relatively stable.

The sufficient sample number in Fig. 16 is smaller than that in common recognition tasks of the visible light images. This is because,

Table 5
Comparison results of different algorithms using 20 positive and 20 negative training samples.

Sequence	INERTIA+MSRC		LBP+MSRC		HOG+MSRC		SDH+SRC		SDH+SVM		HOG+SVM		Proposed	
	PPV (%)	Sensitivity (%)	PPV (%)	Sensitivity (%)	PPV (%)	Sensitivity (%)	PPV (%)	Sensitivity (%)	PPV (%)	Sensitivity (%)	PPV (%)	Sensitivity (%)	PPV (%)	Sensitivity (%)
1	98.75	90.00	100.00	88.75	100.00	91.25	98.75	85.00	98.75	82.50	98.75	96.25	98.75	97.50
2	98.82	85.88	97.65	85.88	98.82	91.76	98.82	77.65	100.00	89.41	98.82	94.12	100.00	100.00
3	98.86	88.64	95.45	81.82	98.86	93.18	98.86	81.82	100.00	75.00	97.73	89.77	98.86	98.86
4	97.89	77.89	96.84	82.11	96.84	90.53	96.84	81.05	96.84	88.42	97.89	95.79	100.00	97.89
5	95.79	80.00	95.79	81.05	98.95	90.53	95.79	84.21	97.89	88.42	98.95	87.37	100.00	96.84
6	97.30	81.08	100.00	81.08	98.65	93.24	95.95	83.78	98.65	83.78	98.65	94.59	98.65	100.00
7	94.32	79.55	98.86	87.50	98.86	84.09	90.91	79.55	98.86	47.73	97.73	84.09	98.86	87.50
8	90.80	45.98	97.70	72.41	98.85	82.76	94.25	80.46	100.00	83.91	98.85	81.61	100.00	79.31
9	100.00	100.00	100.00	95.65	100.00	98.91	100.00	98.91	100.00	97.83	100.00	98.91	100.00	100.00
10	97.50	80.00	98.75	87.50	97.50	86.25	98.75	73.75	100.00	72.50	97.50	88.75	100.00	93.75
Total	96.99	80.90	98.03	84.38	98.73	90.28	96.88	82.75	99.07	81.13	98.50	91.09	99.54	95.14

Table 6
Comparison results of different algorithms using 50 positive and 50 negative training samples.

Sequence	INERTIA+MSRC		LBP+MSRC		HOG+MSRC		SDH+SRC		SDH+SVM		HOG+SVM		Proposed	
	PPV (%)	Sensitivity (%)	PPV (%)	Sensitivity (%)	PPV (%)	Sensitivity (%)	PPV (%)	Sensitivity (%)	PPV (%)	Sensitivity (%)	PPV (%)	Sensitivity (%)	PPV (%)	Sensitivity (%)
1	100.00	92.50	100.00	96.25	100.00	96.25	100.00	78.75	100.00	82.50	100.00	97.50	100.00	97.50
2	98.82	95.29	98.82	95.29	100.00	97.65	100.00	78.82	100.00	97.65	100.00	97.65	100.00	100.00
3	98.86	100.00	98.86	82.95	100.00	90.91	100.00	81.82	100.00	73.87	100.00	89.77	100.00	98.86
4	97.89	74.74	96.84	75.79	97.89	94.74	96.84	82.11	98.95	96.84	98.95	95.79	100.00	98.95
5	98.95	83.16	100.00	85.26	100.00	96.84	98.95	88.42	98.95	100.00	100.00	98.95	98.95	100.00
6	100	97.30	100.00	93.24	100.00	98.65	100.00	85.14	100.00	89.19	100.00	94.59	100.00	100.00
7	97.73	68.18	98.86	87.50	98.86	96.59	93.10	86.21	100.00	85.06	100.00	81.61	100.00	93.10
8	87.36	55.17	97.70	77.01	98.85	87.36	93.10	86.21	100.00	85.06	100.00	81.61	100.00	93.10
9	100.00	100.00	100.00	97.83	100.00	98.91	100.00	100.00	100.00	100.00	100.00	98.91	100.00	100.00
10	97.50	97.50	98.75	87.50	97.50	95.00	98.75	76.25	100.00	72.50	98.75	96.25	100.00	100.00
Total	97.69	86.00	98.96	87.62	99.31	95.25	98.61	84.14	99.77	84.84	99.77	94.79	99.88	98.03

Table 7
Comparison results of different algorithms using 100 positive and 100 negative training samples.

Sequence	INERTIA+MSRC		LBP+MSRC		HOG+MSRC		SDH+SRC		SDH+SVM		HOG+SVM		Proposed	
	PPV (%)	Sensitivity (%)	PPV (%)	Sensitivity (%)	PPV (%)	Sensitivity (%)	PPV (%)	Sensitivity (%)	PPV (%)	Sensitivity (%)	PPV (%)	Sensitivity (%)	PPV (%)	Sensitivity (%)
1	100.00	92.50	100.00	96.25	100.00	96.25	100.00	78.75	100.00	82.50	100.00	96.25	100.00	97.50
2	98.82	92.94	98.82	95.29	100.00	97.65	100.00	87.06	100.00	97.65	100.00	97.65	100.00	100.00
3	98.86	94.32	98.86	82.95	100.00	93.18	100.00	82.95	100.00	75.00	100.00	89.77	100.00	98.86
4	97.89	77.89	97.89	80.00	97.89	96.84	96.84	82.11	98.95	96.84	98.95	95.79	98.95	98.95
5	98.95	84.21	100.00	85.26	100.00	96.84	98.95	89.47	98.95	100.00	100.00	100.00	100.00	100.00
6	100	90.54	100.00	94.59	100.00	98.65	100.00	85.14	100.00	89.19	100.00	94.59	100.00	100.00
7	97.73	77.27	98.86	89.77	98.86	96.59	98.86	81.82	100.00	47.73	100.00	96.59	100.00	92.05
8	88.51	58.62	97.70	83.91	98.85	90.80	93.10	86.21	100.00	85.06	100.00	86.21	100.00	93.10
9	100.00	100.00	100.00	97.83	100.00	98.91	100.00	98.91	100.00	100.00	100.00	100.00	100.00	100.00
10	97.50	97.50	98.75	87.50	97.50	95.00	98.75	76.25	100.00	72.50	98.75	96.25	100.00	100.00
Total	97.80	86.34	99.07	89.12	99.31	96.06	98.61	85.07	99.77	84.95	99.77	95.37	99.88	98.03

for the same classifier, the number of sufficient samples is decided by the target object to be recognized. Without generality, the training samples can be viewed as randomly chosen from the object space. The less information the object contains, the fewer samples it needs, and vice versa. The persons, especially small-size people, in thermal infrared images, usually contain less information than those in the visible light images. Therefore, the number of sufficient samples for the thermal infrared images is not so large.

In the evaluation of pattern recognition, there are various effective methods. Besides PPV and Sensitivity, the ROC curve is also broadly used. Therefore, we summarized the above comparative experimental results with ROC curves, in which the seven methods are compared. As shown in Fig. 17, we can observe that the proposed method outperforms the others.

5. Conclusion

In this paper, a robust method for pedestrian detection in thermal infrared images has been proposed. We adopted a

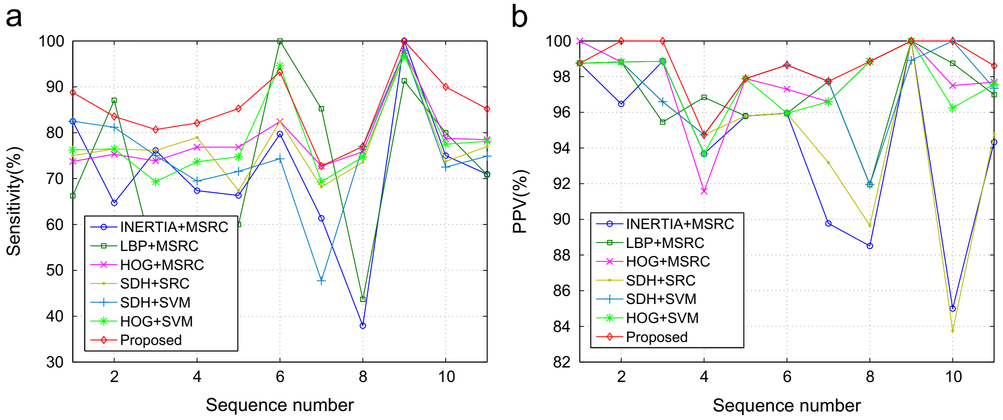

Fig. 12. Comparisons of PPV and Sensitivity rates for different algorithms using 10 training samples (sequence #11 is for the total sequence). (a) The Sensitivity rate. (b) The PPV rate.

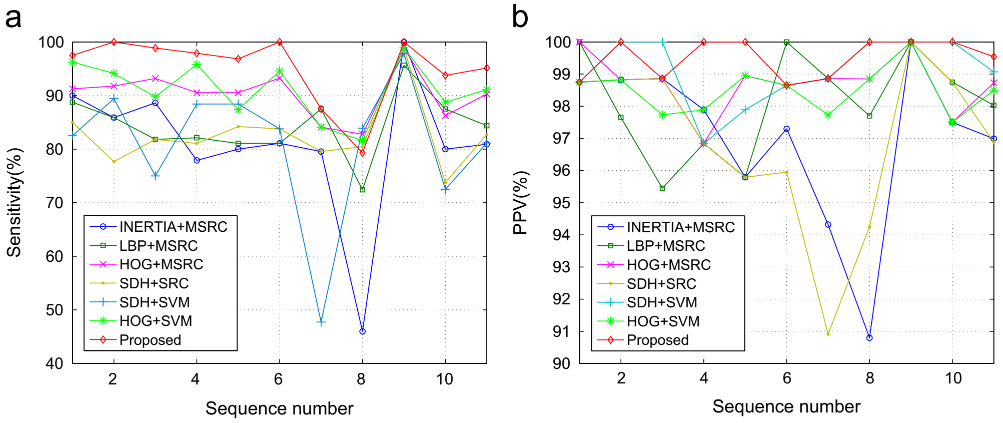

Fig. 13. Comparisons of PPV and Sensitivity rates for different algorithms using 20 training samples (sequence #11 is for the total sequence). (a) The Sensitivity rate. (b) The PPV rate.

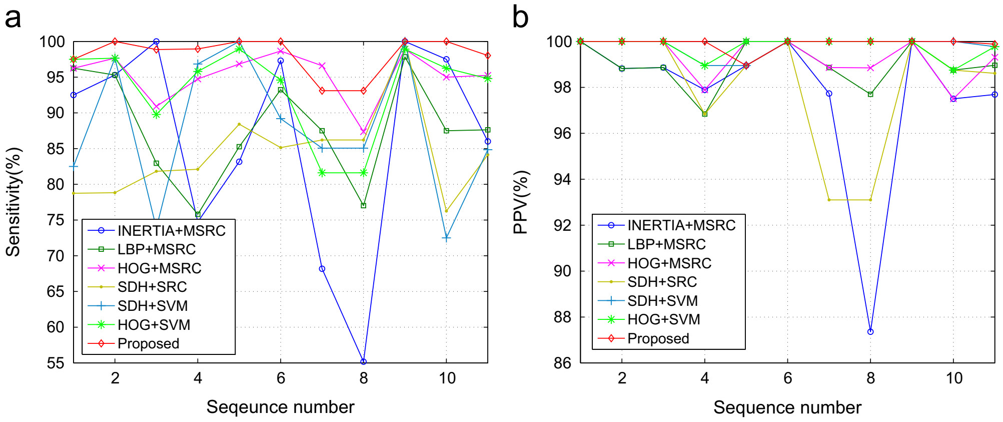

Fig. 14. Comparisons of PPV and Sensitivity rates for different algorithms using 50 training samples (sequence #11 is for the total sequence). (a) The Sensitivity rate. (b) The PPV rate.

discriminative feature, the SDH feature, for representing pedestrians and non-pedestrians. The SDH feature stands for the distribution of Euclidean distances between pairs of randomly selected points from the contour saliency map of an object. This distribution describes the overall shape of the represented object. Furthermore, a robust MSRC is utilized to detect the pedestrians. We modified the SRC classifier so that it can satisfy the requirement of pedestrian classification. The experimental results showed that the proposed method exhibits very high performance by comparing it with other algorithms in different features and classifiers.

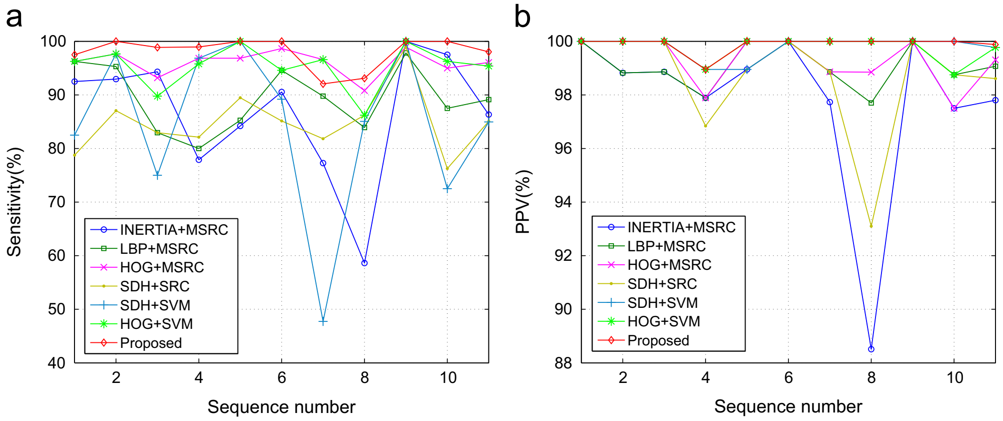

Fig. 15. Comparisons of PPV and Sensitivity rates for different algorithms using 100 training samples (sequence #11 is for the total sequence). (a) The Sensitivity rate. (b) The PPV rate.

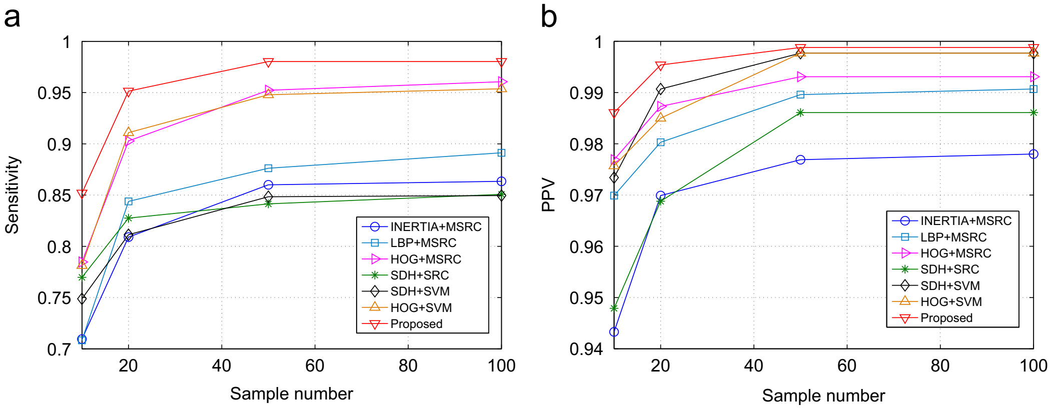

Fig. 16. Comparisons of PPV and Sensitivity rates for different algorithms using 10, 20, 50, 100 training samples. (a) The Sensitivity rate. (b) The PPV rate.

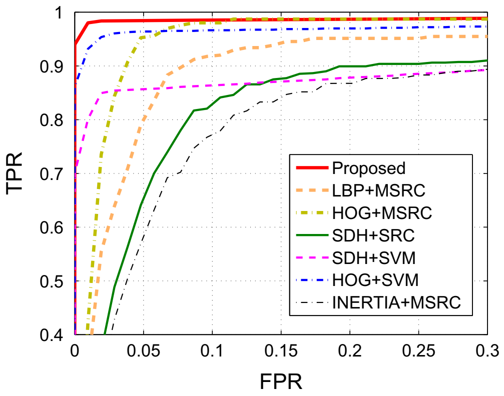

Fig. 17. The ROC curves for the seven algorithms.

Conflict of interest

None declared.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (51405435), Zhejiang Provincial Natural Science Foundation of China (LQ13F030003), the National Program on Key Basic Research Project of China (973 Program) (No. 2011CB706500), and the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry.

References

[1] J.W. Davis, V. Sharma, Background-subtraction using contour-based fusion of thermal and visible imagery, Comput. Vis. Image Understand. 106 (2007) 162–182.
[2] H. Nanda, L. Davis, Probabilistic template based pedestrian detection in infrared videos, in: Intelligent Vehicles Symposium, 2002.
[3] F. Xu, X. Liu, K. Fujimura, Pedestrian detection and tracking with night vision, IEEE Trans. Intell. Transport. Syst. 6 (2005) 1–63.
[4] M. Yasuno, N. Yasuda, M. Aoki, Pedestrian detection and tracking in far infrared images, in: IEEE Intelligent Transportation Systems, 2005.
[5] Y. Owechko, S. Medasani, N. Srinivasa, Classifier swarms for human detection in infrared imagery, in: IEEE International Workshop on Object Tracking and Classification Beyond the Visible Spectrum, 2004.
[6] J. Davis, M. Keck, A two-stage approach to person detection in thermal imagery, in: IEEE OTCBVS WS Series Bench Workshop on Applications of Computer Vision, 2005.
[7] C. Dai, Y. Zheng, X. Li, Pedestrian detection and tracking in infrared imagery using shape and appearance, Comput. Vis. Image Understand. 106 (2007) 288–299.
[8] J. Li, W. Gong, W. Li, X. Liu, Robust pedestrian detection in thermal infrared imagery using the wavelet transform, Infrared Phys. Technol. 53 (2010) 267–273.
[9] J. tao Wang, D. bao Chen, H. yan Chen, J. Yu Yang, On pedestrian detection and tracking in infrared videos, Pattern Recognit. Lett. 33 (2012) 775–785.
[10] B. Ko, D. Kim, J. Nam, Detecting humans using luminance saliency in thermal images, Opt. Lett. 37 (20) (2012) 4350–4352.
[11] S. Agarwal, A. Awan, D. Roth, Learning to detect objects in images via a sparse, part-based representation, IEEE Trans. Pattern Anal. Mach. Intell. 26 (11) (2004) 1475–1490.
[12] K. Toyama, A. Blake, Probabilistic tracking with exemplars in a metric space, Int. J. Comput. Vis. 48 (1) (2002) 9–19.
[13] D. Gavrila, A Bayesian exemplar-based approach to hierarchical shape matching, IEEE Trans. Pattern Anal. Mach. Intell. 29 (8) (2007) 1408–1421.

[14] D. Gavrila, S. Munder, Multi-cue pedestrian detection and tracking from a moving vehicle, Int. J. Comput. Vis. 73 (1) (2007) 41–59.

[15] R. Osada, T. Funkhouser, B. Chazelle, D. Dobkin, Matching 3d models with shape distributions, in: SMI 2001 International Conference on Shape Modeling and Applications, 2001.

[16] J. Wright, A.Y. Yang, A. Ganesh, S.S. Sastry, Y. Ma, Robust face recognition via sparse representation, IEEE Trans. Pattern Anal. Mach. Intell. 31 (2) (2008) 210–227.

[17] M. Bertozzia, A. Broggia, C. Caraffia, M.D. Roseb, M. Felisaa, G. Vezzonia, Pedestrian detection by means of far-infrared stereo vision, Comput. Vis. Image Understand. 106 (2007) 194–204.

[18] J.W. Davis, V. Sharma, Background-subtraction in thermal imagery using contour saliency, Int. J. Comput. Vis. 71 (2) (2006) 161–181.

[19] M. Sezgin, B. Sankur, Survey over image thresholding techniques and quantitative performance evaluation, J. Electron. Imaging 13 (1) (2004) 146–165.

[20] Y. Fang, K. Yamada, Y. Ninomiya, B. Horn, I. Masaki, A shape-independent-method for pedestrian detection with far-infrared-images, IEEE Trans. Vehic. Technol. 53 (5) (2004) 1679–1697.

[21] R. Duda, P. Hart, D. Stork, Pattern Classific., 2nd ed., Wiley, New York, 2000.

[22] S.Z. Li, Face recognition based on nearest linear combinations, in: Proceedings of the IEEE Computer Society Conference on Computer Vision Pattern Recognition, 1998, pp. 839–844.

[23] K.-C. Lee, J. Ho, D. Kriegman, Acquiring linear subspaces for face recognition under variable lighting, IEEE Trans. Pattern Anal. Mach. Intell. 27 (5) (2005) 684–698.

[24] V. Vapnik, The Nature of Statistical Learning Theory, Springer, Berlin.

[25] ⟨http://www.cse.ohio-state.edu/otcbvs-bench/⟩.

[26] M. Heikkilä, M. Pietikäinen, A texture-based method for modeling the background and detecting moving objects, IEEE Trans. Pattern Anal. Mach. Intell. 28 (4) (2006) 657–662.

[27] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: CVPR, 2005, pp. 886–893.

[28] M. Everingham, L.V. Gool, C.K.I. Williams, J. Winn, A. Zisserman, The pascal visual object classes (voc) challenge, Int. J. Comput. Vis. 88 (2010) 303–338.

Xinyue Zhao received the M.S. degree in Mechanical Engineering from Zhejiang University, China in 2008, and the Ph.D. degree in Graduate School of Information Science and Technology from Hokkaido University, Japan, in 2012. She is currently an assistant professor at Zhejiang University, China. Her research interests include computer vision and image processing. She is a member of the IEEE.

Zaixing He received his B.S. and M.S. degrees in Engineering from Zhejiang University, China, in 2006 and 2008, respectively, and then the Ph.D. degree in Graduate School of Information Science and Technology from Hokkaido University, Japan, in 2012. He is currently an assistant professor at Zhejiang University, China. His research interests include sparse recovery, sparse representation, compressed sensing and their applications to image processing, signal processing and pattern recognition. He is a member of the IEEE.

Shuyou Zhang received the M.S. degree in Mechanical Engineering and the Ph.D. degree in State Key Laboratory Of CAD&CG from Zhejiang University, China, in 1991 and 1999, respectively. He is currently a professor at Department of Mechanical Engineering, Zhejiang University, China. He is also the vice administer of Institute of Engineering & Computer Graphics in Zhejiang University, assistant director of Computer Graphics Professional Committee for China Engineering Graphic Society, member of Product Digital Design Professional Committee, and chairman of Zhejiang Engineering Graphic Society. His research interests include product digital design, design and stimulation for complex equipments, and engineering and computer graphics.

Dong Liang received B.E. degree and M.E. degree from Lanzhou University (LZU), China, in 2008 and 2011, respectively. He received Ph.D. from Graduate School of Information Science and Technology, Hokkaido University, Japan, in 2015. He is now an assistant professor in Department of Computer Science & Engineering, Nanjing University of Aeronautics and Astronautics (NUAA), China. His research interests include computer vision and pattern recognition.