Use of Sparse Representation for Pedestrian Detection in Thermal Images

Bin Qi[1,2], Vijay John[1,2], Zheng Liu[1,2], and Seiichi Mita[2]

[1]Intelligent Information Processing Laboratory, Toyota Technological Institute, Nagoya, Japan,
{binqi,vijayjohn,zhengliu}@toyota-ti.ac.jp
[2]Research Centre for Smart Vehicles, Toyota Technological Institute, Nagoya, Japan,
smita@toyota-ti.ac.jp

Abstract

Pedestrian detection plays a paramount role in advanced driver assistant system (ADAS) and autonomous vehicles, especially with the growth of aging population. The purpose of pedestrian detection is to identify and locate people in a dynamic scene or environment. It needs to tackle the challenges such as illumination, color, texture, clothing, and background complexities. Different from visible imaging system, thermal imaging depends on objects' emissivity, and thus has the advantage on discriminating human body from the cool background. In this study, sparse representation is proposed for pedestrian detection in thermal images. Two types of dictionaries, i.e. a generic dictionary optimized by K-SVD and a naive dictionary with basis atoms being directly composed of training samples, are employed to represent image features. In the implementation, a boundary box shrinking scheme is applied to improve the accuracy of the detection through finding proper size for the boundary box. The experimental results demonstrate a comparable performance of the proposed approach.

1. Introduction

Detecting people in images is a basic research of human perception, but plays a fundamental role in a variety of important applications, such as advanced driver assistant system (ADAS), autonomous driving, intelligent video surveillance, victim rescue and people behavior analysis [1, 2, 3, 4]. In ADAS and autonomous driving, pedestrian detection is a key technique to assure the safe driving. Usually, pedestrians are detected from visible color or grayscale images. However, human bodies have the characteristics of rigidity and flexibility, and the appearances can be easily affected by clothing, illumination, occlusion, gesture, visual angle, and complex backgrounds. Pedestrian detection is thus a difficult task with potential challenges for machine perception. Compared with visible images, thermal images are represented with different intensity maps, and not sensitive to illumination change. Besides, thermal images can provide an enhanced spectral range that is imperceptible to human beings and contribute to obvious contrast between objects of high temperature variance and the environment [5, 6, 7, 8, 9]. Some examples are given in Figure 1, where thermal images demonstrate their advantages over visible images in these scenarios. It is possible to detect pedestrians from thermal images with insufficient or over illumination. Moreover, the variability introduced by color, texture, and complex background becomes trivial [7, 9, 10].

Figure 1. Sample visible (top) and thermal (bottom) images.

Pedestrian detection is to determine whether a local image region contains people or not and thus a typical two-class classification problem [1, 11]. The detection procedure can be implemented in two steps: feature extraction and classification. In the learning-based discriminative framework, various feature descriptors and classification approaches have been proposed for use in visible images [1, 2, 12, 13]. Some of the approaches for pedestrian detection from thermal images are based on background modeling and pixel classification. In [14], Davis et al. proposed a background subtraction method, which fused contours from thermal and visible images for persistent pedes-

trian detection. Background subtraction in thermal image identified the regions containing foreground objects. Color information was then used to detect people from the corresponding foreground regions in the visible domain. Feature-based pedestrian detection from thermal images have also been explored. Fang et al. proposed a shape-independent pedestrian detection method in [8]. Pedestrians' horizontal and vertical locations were estimated by a projection-based horizontal segmentation and a brightness/bodyline-based vertical segmentation respectively. In the regions of interest, multidimensional histogram-, inertia-, and contrast-based features were defined and used for further classification. In [9], Li et al. employed a dual-tree complex wavelet transform to decompose regions of interest identified by the high intensity value in a thermal image. Wavelet entropy features were extracted from the high frequency sub-bands. Support vector machine was applied to detect the pedestrian regions. A hybrid-based method was proposed by Zin et al. [5], in which multi-slit feature and HOG (histogram of oriented gradients) feature were fused for pedestrian detection from near infrared images. The feature-level fusion took the advantage of both features on identification and localization of body parts. Image phase congruency feature was used by Olemda et al. [10] to reduce the effects introduced by illumination change in thermal images.

Sparse representation of signals has recently found diverse applications [15, 16, 17]. It computes sparse coefficients with a linear combination of basis atoms from an overcomplete dictionary (i.e. the number of basis atoms exceeds the dimension of signal). Thus, any signal can be represented by more than one combination of different atoms [16]. The purpose of sparse representation is to obtain a compact representation of the observed signal. The selection of the dictionary is critical in such representation [18, 19]. Much of work has been done on how to search the optimal representation with sufficiently sparse or at a required sparsity level [16, 17]. In our work, motivated by the effectiveness of sparse representation based methods for classification problem [19], we exploited the discriminative nature of sparse representation to perform pedestrian detection with two types of dictionaries: a generic dictionary optimized by K-SVD (singular value decomposition) and a naive dictionary with basis atoms directly composed of training samples. The pedestrian detection results are presented by boundary boxes to indicate the location of people. However, it is quite common to assign a large boundary box to a "small" pedestrian in the results. This may lead to the increase of missing rate. Hence, a post-processing scheme, namely boundary box shrinking, is implemented to solve this problem.

The rest of this paper is organized as follows. In section 2, the procedure of pedestrian detection with sparse representation is described. The feature extraction approaches used in this study are presented in section 3. Section 4 explains the boundary box shrinking scheme for post-processing of the detection results. The experimental results can be found in section 5. This paper is concluded in the final section 6.

2. Pedestrian detection via sparse representation

Pedestrian detection uses samples from foreground (with pedestrian) and background (without pedestrian) to train a classifier. Each sample is represented with certain feature vector of length d as an input to the classifier. The m foreground and n background training samples can be organized in the columns of two matrices $F = [f_1, f_2, ..., f_m] \in \mathbb{R}^{d \times m}$ and $B = [b_1, b_2, ..., b_n] \in \mathbb{R}^{d \times n}$ respectively.

2.1. Learning dictionary for sparse representation

2.1.1 Generic dictionary optimized by K-SVD

In spare representation, natural samples can be represented as a linear combination of basis atoms with associated sparse coefficients. With a dictionary matrix $D \in \mathbb{R}^{d \times K}$ that contains K basis atoms $\{a_i\}_{i=1}^{K}$ at each column, a sample $y \in \mathbb{R}^{d \times 1}$ can be represented as a sparse linear combination of these atoms. The K-SVD algorithm is an efficient technique for learning the dictionary from given training samples. Generally, the K-SVD algorithm is initialized with a dictionary D_0, the number of iteration, and a set of training samples arranged as the columns of matrix T. The algorithm aims to iteratively update the dictionary to achieve the best possible sparse representations of the samples in T with strict sparsity constraints, by solving the following optimization problem [16, 17]:

$$\{D, X\} = \underset{D,X}{\operatorname{argmin}} \|T - DX\|_2 \quad \forall i, \|x_i\|_0 \leq L \tag{1}$$

The i-th column of X, x_i, is the sparse coefficient of the i-th column of T. Operator $\|\cdot\|$ is the l^0 pseudo-norm which counts the non-zero entries and L is the sparsity level. The K-SVD algorithm alternates between sparse coding of the training samples with respect to the current dictionary and an update process for the dictionary atoms so as to better fit the training samples [18]. Given the training sample matrices F and B, the associated foreground and background dictionary D_f and D_b can be obtained by solving the following optimization problem:

$$\begin{cases} D_f = \underset{D}{\operatorname{argmin}} \|F - DX_f\|_2 \quad \forall i, \|x_{f,i}\|_0 \leq L \\ D_b = \underset{D}{\operatorname{argmin}} \|B - DX_b\|_2 \quad \forall i, \|x_{b,i}\|_0 \leq L \end{cases} \tag{2}$$

where X_f and X_b are the sparse coefficients of foreground and background training samples respectively while the i-

th column of X_f and X_b are $x_{f,i}$ and $x_{b,i}$. And L is the sparsity level.

2.1.2 Represent test sample as a sparse linear combination of training samples

In sparse representation, the input sample should be sparsely represented by the linear combinations of basis atoms. It seeks to minimize the difference between input sample and associated spare representations. Although the generic dictionary has a good performance on reconstructing the input samples, for pedestrian detection, we only care either the foreground or background dictionary can fit the testing sample better. Thus, the testing sample can be reconstructed or represented directly with a dictionary where the basis atoms come from the training samples. The difference between the input sample and corresponding representation can be minimized when large number of training samples are available.

2.2. Detection based on sparse representation

Given a new test sample y, its sparse representation x_f and x_b are computed with foreground and background dictionary D_f and D_b respectively. With each dictionary, one can approximate the given test sample y as $\hat{y}_f = D_f x_f$ and $\hat{y}_b = D_b x_b$. Thus, the test sample y can be classified based on these approximations by assigning it to the class that minimizes the residual $r(y)$ between y and $\hat{y}$:

$$\underset{j}{\operatorname{argmin}} r_j(y) = \underset{j}{\operatorname{argmin}} \|y - D_j x_j\|_2 \quad j = f, b \tag{3}$$

The procedures of pedestrian detection with the two types of dictionaries are further detailed and summarized in following Algorithm 1 and 2. The use of the generic dictionary optimized by K-SVD is described Algorithm 1. Algorithm 2 summarizes the procedure with the naive dictionary composed of training samples.

Algorithm 1 Sparse representation based classification with generic dictionary optimizing by K-SVD (SRC-K-SVD)

Input: matrix of foreground training samples $F \in \mathbb{R}^{d \times m}$; matrix of background training samples $B \in \mathbb{R}^{d \times n}$; test sample $y \in \mathbb{R}^{d \times 1}$.

Output: identify(y) and score(y).

1: Compute foreground dictionary D_f and background dictionary D_b by solving Eq.(2);

2: Solve the following l^0-minimization problem:
$$\begin{cases} x_f = \underset{x}{\operatorname{argmin}} \|y - D_f x\|_2, & \|x\|_0 \le L \\ x_b = \underset{x}{\operatorname{argmin}} \|y - D_b x\|_2, & \|x\|_0 \le L \end{cases}$$

3: Compute the residuals $\begin{cases} r_f(y) = \|y - D_f x_f\|_2 \\ r_b(y) = \|y - D_b x_b\|_2 \end{cases}$

4: Output: $\begin{cases} \text{identify}(y) = \underset{j}{\operatorname{argmin}}\, r_j(y),\ j = f, b \\ \text{score}(y) = |r_b(y) - r_f(y)| \end{cases}$

Algorithm 2 Sparse representation based classification with the naive dictionary composed of training samples (SRC-TS)

Input: matrix of foreground training samples $F \in \mathbb{R}^{d \times m}$; matrix of background training samples $B \in \mathbb{R}^{d \times n}$; test sample $y \in \mathbb{R}^{d \times 1}$.

Output: identify(y) and score(y).

1: Set foreground and background dictionary as $D_f = F$, $D_b = B$;

2: Solve the following l^0-minimization problem;
$$\begin{cases} x_f = \underset{x}{\operatorname{argmin}} \|y - D_f x\|_2, & \|x\|_0 \le L \\ x_b = \underset{x}{\operatorname{argmin}} \|y - D_b x\|_2, & \|x\|_0 \le L \end{cases}$$

3: Compute the residuals $\begin{cases} r_f(y) = \|y - D_f x_f\|_2 \\ r_b(y) = \|y - D_b x_b\|_2 \end{cases}$

4: Output: $\begin{cases} \text{identify}(y) = \underset{j}{\operatorname{argmin}}\, r_j(y),\ j = f, b \\ \text{score}(y) = |r_b(y) - r_f(y)| \end{cases}$

3. Feature extraction approaches

Numerous approaches have been proposed for the image feature extraction in pedestrian detection [1]. However, their performances on thermal images have not been fully explored. In this study, we considered three approaches, i.e. HOG (histogram of oriented gradients), HPC (histogram of phase congruency), and HSC (histogram of sparse code), for pedestrian detection [20, 10, 21]. These approaches are briefly described below.

3.1. Histogram of oriented gradients

In the implementation of HOG, image local gradients are binned according to their orientations, weighted by the magnitudes [20]. Within a spatial grid of cells in block, a feature vector is extracted by sampling the histograms from the contributing spatial cells. The feature vectors for all blocks are concatenated to yield a final feature vector. Detailed information is available in [20] and will not be duplicated herein.

3.2. Histogram of phase congruency

To describe the significance of image features, people are looking for the measures that are invariant with respect to image illumination and magnification change. A feature perception model, which postulates that features are perceived at points in the image where Fourier components were maximally in phase, was proposed [22, 23]. The phase

congruency function is defined in terms of Fourier series expansion of a signal at some location x as [23]:

$$PC(x) = \max_{\theta \in [0,2\pi]} \frac{\sum_n A_n \cos(\phi_n(x) - \theta)}{\sum_n A_n} \quad (4)$$

where A_n is the amplitude of the n-th Fourier component; $\phi_n(x)$ is the local phase of Fourier component at position x; The value θ that maximizes Eq.(4) is the amplitude weighted mean local phase angle of all the Fourier terms at the point being considered [23]. For using phase congruency for images, Kovesi proposed a scheme to calculate the phase congruency with logarithmic Gabor wavelets, which allow arbitrarily large bandwidth filters to be constructed while still maintaining a zero DC component in the even-symmetric filter. The 2D Phase congruency is given by the summation over orientation o and scale n [23]:

$$PC(x) = \frac{\sum_o \sum_n W_o(x) \lfloor A_{no}(x) \Delta\Phi_{no}(x) - T_o \rfloor}{\sum_o \sum_n A_{no}(x) + \sigma} \quad (5)$$

where $\lfloor \cdot \rfloor$ is the floor function that enclosed quantity is not permitted to be negative. A_n represents the amplitude of the n-th component in the Fourier series expansion. A small positive constant σ is added to the denominator in case of small Fourier amplitudes. T_o compensates for the influence of noise and is estimated empirically. $\Delta\Phi_{no}(x)$ is the phase deviation.

Phase congruency has been applied to varied applications including pedestrian detection [10, 24]. The outputs of the calculation contains a phase congruency map and an orientation map corresponding to the maximum features. Similar to HOG, the extracted feature is represented as the histogram of phase congruency (HPC) for further use.

3.3. Histogram of sparse codes

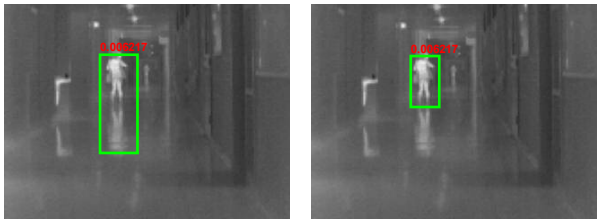
The histogram of sparse codes (HSC) is based on the sparse coding technique, where each local patch of an image is represented with a sparse set of codewords [21]. Given a set of input images, each image is split into small patches. These patches are then arranged as columns of matrix Y_i. The set of input images can be represented as $Y = \{Y_i\}_{i=1}^n$. The K-SVD algorithm is used to jointly find a dictionary D and a set of associated sparse code matrixes $X = \{X_i\}_{i=1}^n$ by minimizing the following optimization problem:

$$D = \underset{D}{\operatorname{argmin}} \sum_{i=1}^{n} \|Y_i - DX_i\|_2 \quad \forall i, j, \ \|x_{i,j}\|_0 \leq L \quad (6)$$

where $x_{i,j}$ is j-th column of X_i, which is the sparse code computed from the j-th patch of Y_i. Let x be the sparse code computed at an unknown patch y. For each non-zero entry x_i in x, soft binning is used to assign its absolute $|x_i|$ to one of the four spatially surrounding cells according to their orientations indicated by i. Within a spatial grid of cells in a block, a feature vector is extracted by sampling the histograms from the contributing spatial cells. The feature vector for all blocks are concatenated to yield a final feature vector, which is called histogram of sparse codes.

4. Boundary box shrinking

The results of pedestrian detection are presented by boundary boxes to indicate the detected pedestrians in the image. However, image noises and modeling errors may cause small pedestrians associated with large boundary boxes (See Figure 2). This may lead to a higher missing rate. Motivated by the idea of local outlier factor (LOF) [25], a boundary box shrinking (BBS) scheme is implemented to post process the detection results.



(a) (b)

Figure 2. Example of boundary box shrinking. (a) Original detection result, (b) Detection result after boundary box shrinking.

Each pedestrian pixel is assigned a degree of being an outlier with respect to the surrounding neighbors. The margin regions that contain outliers are removed. To describe the BBS scheme, some relevant definitions are given below.

Definition 1 (k^{th}-distance)**.** *Given a data set S, for any positive integer k, the k^{th}-distance of pixel p, denoted as $dis_k(p)$, is defined as the distance $d(p,q)$ between p and a pixel $q \in S$ such that*

(1) at least k pixels $q' \in S \backslash \{p\}$ satisfy
$$d(p,q') \leq d(p,q)$$

(2) at most $k-1$ pixels $q' \in S \backslash \{p\}$ satisfy
$$d(p,q') < d(p,q)$$

Definition 2 (k^{th}-distance neighbor)**.** *Given the k^{th}-distance of pixel p, the k^{th}-distance neighbor of p, denoted as $N_{k-dis}(p)$, is one of the pixels whose distance from p is equal to $dis_k(p)$. If several pixels satisfy $d(p,q) = dis_k(p)$, a random pixel q is selected as the k^{th}-distance neighbor of p.*

Definition 3 (outlier degree)**.** *The outlier degree of a pixel p is defined as the ratio between k^{th}-distance of p and k^{th}-distance of $N_{k-dis}(p)$, which is given by*

$$OD(p) = \frac{dis_k(p)}{dis_k(N_{k-dis}(p))} \quad (7)$$

The possibility for a pixel p to be an outlier depends on the value of outlier degree. The margin regions that contain such kind of pixels are useless for detection. The BBS process is applied along four orientations: top-to-bottom, bottom-to-top, left-to-right, and right-to-left. The top-to-bottom process is illustrated with a flowchart in Figure 3 as an example.

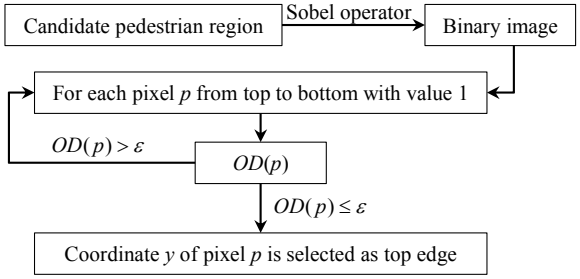


Figure 3. Processing procedure for boundary box shrinking (top-to-bottom)

For the candidate pedestrian regions, Sobel operator is applied to the original thermal images to get an binary edge map. Ideally, the pedestrians' contours have value 1 in the edge map which have many nearby neighbors. Isolated points that locate far from the surrounding neighbors are regarded as outliers or background pixels. Thus, corresponding regions will be removed from the detection results. In this procedure, the outlier degree will assess whether a pixel is an outlier or not, where ε is a loosely selected parameter. Whenever a pixel p satisfied $OD(p) \leq \varepsilon$, it reaches the top edge of the boundary box. Corresponding vertical coordinate of the associated pixel p is selected as the top edge of the boundary box.

5. Experimental results

5.1. Dataset

The dataset used in the experiments contains 2084 thermal images [26]. The ground truth data manually marked from three scenes is used to evaluate the detection results. This dataset contains pedestrians with various poses from front, back or flank views at different scales. The training data contains 2000 pedestrians and 2000 non-pedestrian samples. Figure 4 shows a few examples of pedestrian and non-pedestrian samples.

5.2. Evaluation method

The evaluation was carried out by using a bounding box (BB) and a score or confidence value for the detection result. The non-maximal suppression (NMS) operation is applied to the detection results to get the final output. The evaluation is then be performed on the final output bounding boxes. If the overlap ratio between a detected BB (BB_{dt}) and a ground truth BB (BB_{gt}) is large enough, a potential

Figure 4. Pedestrian (top) and non-pedestrian (bottom) training samples in the dataset.

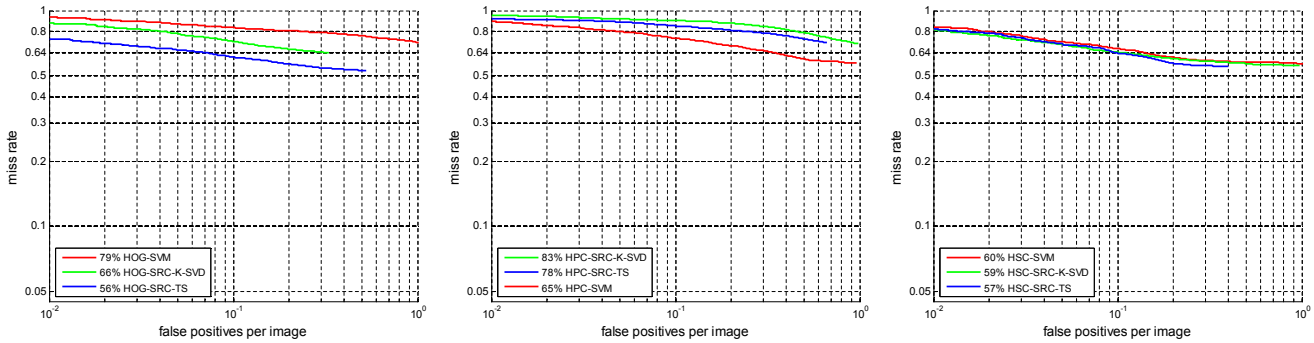
match is achieved. In the experiments, an overlap ratio of 0.5 is selected in Eq.(8).

$$\frac{\text{area}(\mathrm{BB}_{dt} \cap \mathrm{BB}_{gt})}{\text{area}(\mathrm{BB}_{dt} \cup \mathrm{BB}_{gt})} > \text{overlap ratio} \tag{8}$$

The matching between BB_{dt} and BB_{gt} is performed greedily. The matching process is ordered by the confidence of each BB_{dt} so that the BB_{dt} with the highest confidence will be matched first. Sometimes, a BB_{dt} could match multiple BB_{gt}. In this case, the match with the highest overlap ratio will be selected. Whenever a BB_{gt} is matched with a BB_{dt}, other BB_{dt} with lower confidence are not allowed to match this BB_{gt}. That is, each BB_{dt} and BB_{gt} may be matched at most once. Unmatched BB_{dt} are counted as false positives and unmatched BB_{gt} are counted as false negatives. To compare the performance of each detector, the miss rate against false positives per image (FPPI) is plotted with log-log plots. For visual representation, the log-average miss rate for each detector is given, which is computed by averaging the miss rate at nine FPPI rates.

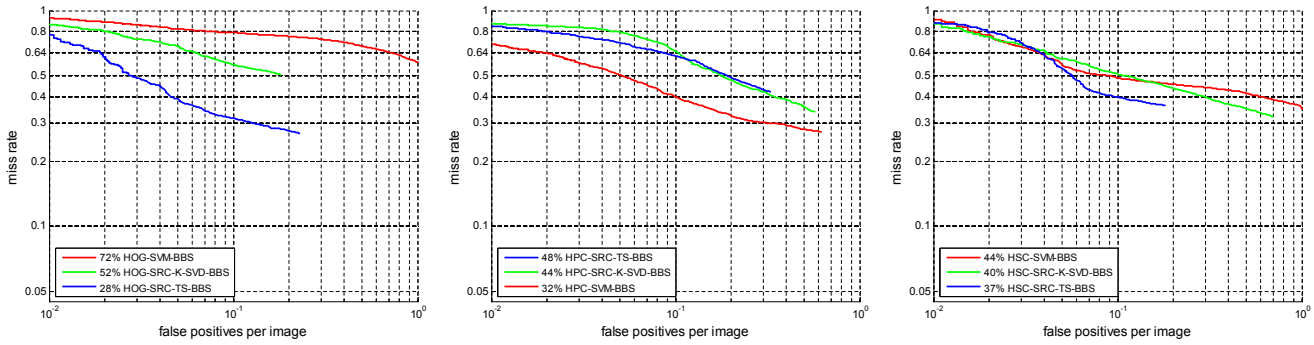
5.3. Results

In the experiments, support vector machine is selected as a contrastive classifier. Figure 5 shows the detection results achieved by difference combination of the three feature extraction methods and classification methods, i.e. SVM, SRC-K-SVD, and SRC-TS. Miss rate versus false positives per image is plotted and log-average miss rate is employed as a common reference value to evaluate the performance. For HOG and HSC feature, SRC-TS has a better performance, i.e. with a lower log-average miss rate across the false positives per image. The HOG feature performed relatively worse especially with the SVM classifier. The SVM achieved a 65% log-average miss rate with HPC, which is better than both SRC-K-SVD and SRC-TS. However, among all the combinations of feature extraction approaches and classifiers, HOG-SRC-TS demonstrates the best performance. Figure 6 shows the detection performance with boundary box shrinking. After post process with boundary box shrinking, all the performance is improved by the decreases of the log-average miss rate: SVM 18.7%, SRC-K-SVD 24%, and SRC-TS 26%. Fig-



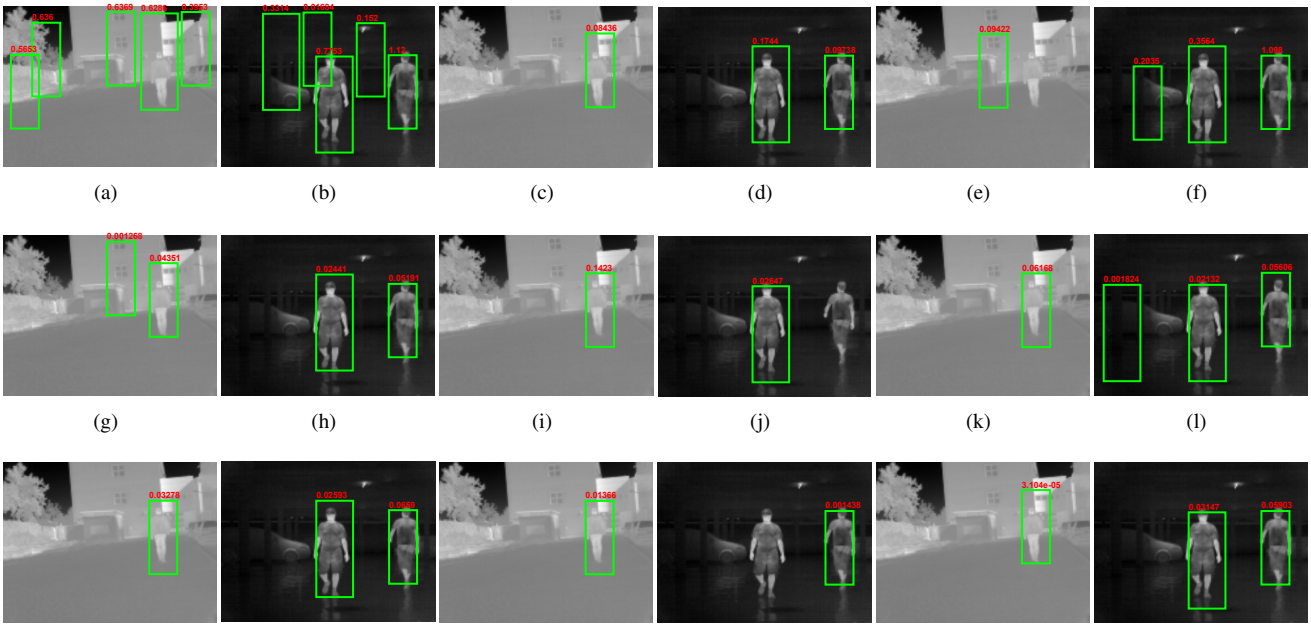
(a) (b) (c)

Figure 5. Evaluation of different classifiers with feature extraction approaches. (a) HOG, (b) HPC, (c) HSC.



(a) (b) (c)

Figure 6. Evaluation of different classifiers with feature extraction approaches after boundary box shrinking. (a) HOG, (b) HPC, (c) HSC.



(a) (b) (c) (d) (e) (f)

(g) (h) (i) (j) (k) (l)

(m) (n) (o) (p) (q) (r)

Figure 7. Pedestrian detection results from different scenes. The images are presented in the format: (scene 1, scene 2) algorithm. (a, b) HOG-SVM, (c, d) HPC-SVM, (e, f) HSC-SVM, (g, h) HOG-SRC-K-SVD, (i, j) HPC-SRC-K-SVD, (k, l) HSC-SRC-K-SVD, (m, n) HOG-SRC-TS,(o, p) HPC-SRC-TS, and (q, r) HSC-SRC-TS.

ure 7 gives the examples of the detection results in different scenes.

6. Conclusion

In this paper, sparse representation based classification (SRC-K-SVD and SRC-TS) and a post-processing approach (boundary box shrinking) are proposed for pedestrian detection. The sparse coefficients are computed with two types of dictionaries. The input image can be represented by a linear combination of basis atoms from the dictionary. The detection is accomplished by assign the input image to the class which minimizes the residual between this image and corresponding approximation. In the experiments, three kinds of feature extraction approaches were employed. The detection with sparse representation together with the HOG feature demonstrated the best performance. And the boundary box shrinking scheme contributes to the decrease of the log-average miss rate.

References

[1] P. Dollar, C. Wojek, B. Schiele, and P. Perona. Pedestrian detection: An evaluation of the state of the art. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34:743–761, 2012.

[2] M. Enzweiler and D. M. Gavrila. Monocular pedestrian detection: Survey and experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31:2179–2195, 2009.

[3] S. Munder and D. M. Gavrila. An experimental study on pedestrian classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28:1863–1868, 2006.

[4] S. Agarwal, A. Awan, and D. Roth. Learning to detect objects in images via a sparse, part-based representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26:1475–1490, 2004.

[5] T. T. Zin, P. Tin, and H. Hama. Pedestrian detection based on hybrid features using near infrared images. *International Journal of Innovative Computing, Information and Control*, 7:5015–5025, 2011.

[6] Q. Liu, J. Zhuang, and J. Ma. Robust and fast pedestrian detection method for far-infrared automotive driving assistance systems. *Infrared Physics and Technology*, 60:288–299, 2013.

[7] M. Bertozzi, A. Broggi, C. Caraffi, M. Del Rose, M. Felisa, and G. Vezzoni. Pedestrian detection by means of far-infrared stereo vision. *Computer Vision and Image Understanding*, 106:194–204, 2007.

[8] Y. Fang, K. Yamada, Y. Ninomiya, B. K. Horn, and I. Masaki. A shape-independent method for pedestrian detection with far-infrared images. *IEEE Transactions on Vehicular Technology*, 53:1679–1697, 2004.

[9] J. Li, W. Gong, W. Li, and X. Liu. Robust pedestrian detection in thermal infrared imagery using the wavelet transform. *Infrared Physics and Technology*, 53:267–273, 2010.

[10] D. Olmeda, A. Escalera, and J. M. Armingol. Contrast invariant features for human detection in far infrared images. In *IEEE Intellignet Vehicles Symposium*, 2012.

[11] K. Goto, K. Kidono, Y. Kimura, and T. Naito. Pedestrian detection and direction estimation by cascade detector with multi-classifiers utilizing feature interaction descriptor. In *IEEE Intelligent Vehicles Symposium*, 2011.

[12] J. F. Ge, Y. P. Luo, and G. M. Tei. Real-time pedestian detection and tracking at nighttime for driver-assistance systems. *IEEE Transactions on Intelligent Transportation Systems*, 10:283–298, 2009.

[13] D. Geronimo, A. M. Lopez, A. D. Sappa, and T. Graf. Survey of pedestrian detection for advanced driver assistance systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32:1239–1258, 2010.

[14] J. W. Davis and V. Sharma. Background-subtraction using contour-based fusion of thermal and visible imagery. *Computer vision and image understanding*, 106:162–182, 2007.

[15] E. Elhamifar and R. Vidal. Sparse subspace clustering: Algorithm, theory, and application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35:2765–2781, 2013.

[16] M. Aharon, M. Elad, and A. Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing*, 54:4311–4322, 2006.

[17] M. Elad and M. Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image Processing*, 15:3736–3745, 2006.

[18] L. Kang, C. Hsu, H. Chen, C. Lu, C. Lin, and S. Pei. Feature-based sparse representation for image similarity assessment. *IEEE Transactions on Multimedia*, 13:1019–1030, 2011.

[19] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31:210–227, 2009.

[20] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *IEEE Computer Vision and Pattern Recognition*, 2005.

[21] X. Ren and D. Ramanan. Histograms of sparse codes for object detection. In *IEEE Computer Vision and Pattern Recognition*, 2013.

[22] M. C. Morrone and R. A. Owens. Feature detection from local energy. *Pattern Recognition Letters*, 6:303–313, 1987.

[23] P. Kovesi. Phase congruency: A low-level image invariant. *Psychological Research*, 64:134–148, 2000.

[24] Z. Liu and R. Laganiere. Phase congruence measurement for image similarity assessment. *Pattern Recognition Letters*, 28:166–172, 2007.

[25] M. M. Breunig, H. P. Kriegel, R. T. Ng, and J. Sander. Lof: Identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, 2000.

[26] D. Olmeda, C. Premebida, U. Nunes, J. Armingol, and A. Escalera. LSI far infrared pedestrian dataset. http://e-archivo.uc3m.es/handle/10016/17370.