

RESEARCH ARTICLE

# Local structure preserving sparse coding for infrared target recognition

Jing Han, Jiang Yue, Yi Zhang, Lianfa Bai\*

Jiangsu Key Laboratory of Spectral Imaging and Intelligent Sense, Nanjing University of Science and Technology, Nanjing, Jiangsu, China

\* blf@njust.edu.cn



## Abstract

Sparse coding performs well in image classification. However, robust target recognition requires a lot of comprehensive template images and the sparse learning process is complex. We incorporate sparsity into a template matching concept to construct a local sparse structure matching (LSSM) model for general infrared target recognition. A local structure preserving sparse coding (LSPSc) formulation is proposed to simultaneously preserve the local sparse and structural information of objects. By adding a spatial local structure constraint into the classical sparse coding algorithm, LSPSc can improve the stability of sparse representation for targets and inhibit background interference in infrared images. Furthermore, a kernel LSPSc (K-LSPSc) formulation is proposed, which extends LSPSc to the kernel space to weaken the influence of the linear structure constraint in nonlinear natural data. Because of the anti-interference and fault-tolerant capabilities, both LSPSc- and K-LSPSc-based LSSM can implement target identification based on a simple template set, which just needs several images containing enough local sparse structures to learn a sufficient sparse structure dictionary of a target class. Specifically, this LSSM approach has stable performance in the target detection with scene, shape and occlusions variations. High performance is demonstrated on several datasets, indicating robust infrared target recognition in diverse environments and imaging conditions.

## OPEN ACCESS

**Citation:** Han J, Yue J, Zhang Y, Bai L (2017) Local structure preserving sparse coding for infrared target recognition. PLoS ONE 12(3): e0173613. <https://doi.org/10.1371/journal.pone.0173613>

**Editor:** Zhao Zhang, Soochow University, CHINA

**Received:** August 25, 2016

**Accepted:** February 22, 2017

**Published:** March 21, 2017

**Copyright:** © 2017 Han et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All relevant data are within the paper and its Supporting Information files.

**Funding:** This work was supported by the Natural Science Foundation of China (61231014 L.B., 61501235 J.H.) and the Fundamental Research Funds for the Central Universities (30915011335 J.Y., 30916011320 J.H.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

## Introduction

Automatic recognition of targets in infrared images is a challenging problem because of some inherent characteristics of the infrared image itself. Infrared images have the vignetting effect and smooth texture, contain different level of noise and pixel mixing. Infrared targets have inconsistent brightness, which is related to target orientation, surface material, etc. [1]. Besides, non-rigid targets have diverse postures, shapes, and sizes, such as humans and animals [2]. Combined with the imaging angles, scene clutters, background occlusions, and other factors [2–4], these could all be the constraints of infrared target recognition. Therefore, achieving robust target recognition with anti-interference ability (noise, fuzzification, occlusion, target shape, and scene changes) from an infrared image is still a challenging work.

Sparse representation classifier (SRC) is well applied to image classification and target recognition [5–7], which is insensitive to noise and missing data [5]. Based on SRC, JEDL obtains

a simple and efficient sparse codes auto-extractor and a linear multi-class classifier from one objective function by minimizing the sparse reconstruction, discriminative sparse-code, code approximation and classification errors simultaneously [8]. But the training based methods need large samples for non-rigid targets. Indeed, there studied a series of unsupervised feature selection methods. CGSSL framework exploits the cluster analysis and structural analysis with sparsity simultaneously [9]; NSCR algorithm exploits nonnegative cluster analysis and redundancy control with row sparsity simultaneously [10]; RSSL algorithm exploits the intrinsic geometric structure of data, and the local and global structural consistencies over labels simultaneously [11]. They utilize the cluster analysis to realize a series of unsupervised feature selection methods, and belong to the re-constraint of the cluster process essentially, by adding sparsity and correlation constrains to the cluster transformation matrix. The ability of feature representation depends to some extent on the cluster effect, and there may be over-fitting of false clusters, so that the actual separability of the low dimensional features is affected.

For the practical application of infrared target detection, we study the method based on small sample or limited sample matching, which requires the sparse coding (Sc) to represent the intrinsic features of infrared targets accurately and stably. IR image has blurred texture details, pixel mixing and is affected by noise. It is necessary to study a new measure for stability analysis of the IR targets. Actually, in conditions of complex noise, blurred details and diverse imaging environments, the local essential structures of pure signals are invariable, both in sharp transitions and texture areas. We can exploit the stable prior of local spatial structure to suppress the noise and fuzzification problem, which can enhance the robustness of sparse representation in IR image. On the other hand, the local spatial structure can reflect the texture details, it can be combined with sparsity analysis to perform a more discriminative feature selection of natural IR images.

Similarly, JELSR integrates the merits of both manifold learning and sparse regression, and proposes an efficient method for unsupervised feature selection [12]. SDPE and SPPE preserving pairwise similarities between data points in addition to preserving the sparse characteristics [13]. In essence, these algorithms add the constraints of sparse mapping and classification in the objective function, or add different constraints of transformation matrix, like sparsity, on the basis of linear manifold dimension reduction methods. However, the sparseness here is mainly used to recover the noise reduction data of the original sample, which makes the essential features of dimensionality reduction more accurate.

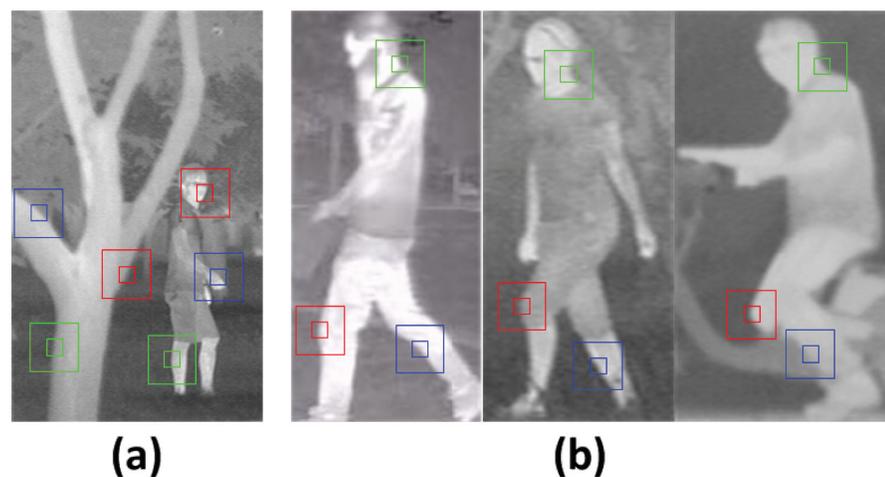
In fact, sparsity cannot reflect the locality. As suggested in [14,15], locality is more essential than sparsity, as locality must lead to sparsity but not necessarily vice versa. However, existing local sparse methods mainly deal with the sparsity of local features in image [14] or constrain the sparsity on local dictionary bases [15]. The latter is the structured sparse, which executes the sparse constraint on group dictionary bases to solve the global similarity preserving problem smoothly [16–19]. Local dictionary constraints can enhance the stability and separability of sparse representations, but they also focus on the global feature selection. Similarly, Gao proposed a laplacian sparse coding (LSc) [20]. By adding a similarity preserving term to the formulation of classical sparse coding, LSc can preserve the global similarity of features in the sparse coding process. Peng establishes a discriminative regression approach (DR) by explicitly incorporating the discriminative information into regression in the instance space and the coefficient space jointly for high-dimensional and large-scale data [21]. We include the constraint of spatial local consistency in sparse representation process, and design an effective algorithm, to solve the corresponding optimization problem, and construct a framework to apply in the infrared object detection.

For target detection analyses, we consider to learn the local areas of original samples. Though the whole structure of non-rigid samples has great differences, the characteristics in

their local region are consistent. The sparse solution can provide the consistency between the test signal and training samples [22]. We divide the integral sample into small patches, the small subsamples have strong sparsity and big subsamples contain more structure information. As shown in Fig 1(A), part of patches in the trunk have similar features to the human body (small windows with same color in Fig 1(A)), which may cause the consistent sparse representations. There are quite different structures between bigger subregions in the trunk and the human body (big windows with same color in Fig 1(A)). But the various structures in big regions may cause weak sparsity. Therefore, we propose a local structure preserving sparse coding (LSPSc) algorithm to combine the sparsity and structure effectively, by adding a constraint of spatial local manifold among samples into Sc. LSPSc considers both the sparsity of small patches and the structure in big subregions, which ensures both the sparsity and locality (local area constituted by the patch and its neighborhood). Different from the above algorithms, LSPSc embeds manifold structure constraints into robust sparse representation, and the solution process of corresponding optimization problem is different.

The kernel trick can capture the nonlinear relationship from features, which may reduce the quantization error and boost the sparse coding performance. It has been pointed out that sparsity of coefficient provides no real help for the accuracy in small sample classification [23,24]. To solve this problem, Gao successfully combined kernel techniques with sparse representation [25,26] and incorporated kernel sparse coding (KSc) into a spatial pyramid matching (SPM) algorithm, which is successfully applied in image classification. Kernel sparse representation-based classification (KSRC) is then proposed [27–29]. KSRC maps the nonlinearly separable features into a high-dimensional kernel space and performs in the new space, in which complex inherent structural differences are more easily grouped together and linearly separable. SRC in kernel space can be formulated in terms of the inner products [30,31]. Cheng aims at the high-dimensional data to construct a minimax framework for multiclass classification, which can account for nonlinearity in the input space by using kernel techniques [32]. Motivated by these, we further extend LSPSc to kernel LSPSc (K-LSPSc) to adapt general nonlinear data, and design the solution procedure of optimization problem in kernel space.

Based on LSPSc and K-LSPSc, we construct a general IR target recognition method, local sparse structure matching (LSSM). Combining the discriminative feature, local matching and statistics idea, LSSM can use several template images to achieve the robust detection of IR



**Fig 1. Analysis of the spatial local sparsity and structure in human targets.**

<https://doi.org/10.1371/journal.pone.0173613.g001>

targets. When there are occlusions, shape and scene changes, part of the local sparse structures of the target remain unchanged (Fig 1(B)), only the target area having consistent basic elements and similar structures with the template can obtain a low objective. Consequently, LSSM only requires the template set to have enough local structure and basic information of an object class, which is named as “simple template set” later. It has weak dependence on the comprehensive template images. Compared with the clustering and classification issues in [5,8,20,21], LSSM focuses on the object detection framework more, and different from the entire template matching in [33–38], LSSM is a local analysis method.

The paper is organized as follows: Section 2 details the formulation and implementation of LSPSc and K-LSPSc, presents the LSSM and the analysis of LSPSc/K-LSPSc in general infrared target recognition. In Section 3, we demonstrate the performance of LSSM with some experimental results on both visible and infrared image datasets. Finally, we conclude the paper in Section 4.

## Methods and models

### Sparse coding

Traditional sparse coding considers the maximum a posteriori estimate of the basis and coefficients, assuming a uniform prior on the basis, by solving the solution to the following optimization problem [39]:

$$\min_{\mathbf{B}, \mathbf{S}} \sum_i \|\mathbf{x}_i - \mathbf{B}\mathbf{s}_i\|^2 + \lambda \sum_i \|\mathbf{s}_i\|_1, \quad s.t. \|\mathbf{b}_i\|^2 \leq c, \quad i = 1, 2, \dots, M \quad (1)$$

Denote  $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N] \in R^{P \times N}$  as all the samples,  $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_M] \in R^{P \times M}$  as the dictionary matrix, and  $\mathbf{S} = [\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_N] \in R^{M \times N}$  as the coefficient matrix.

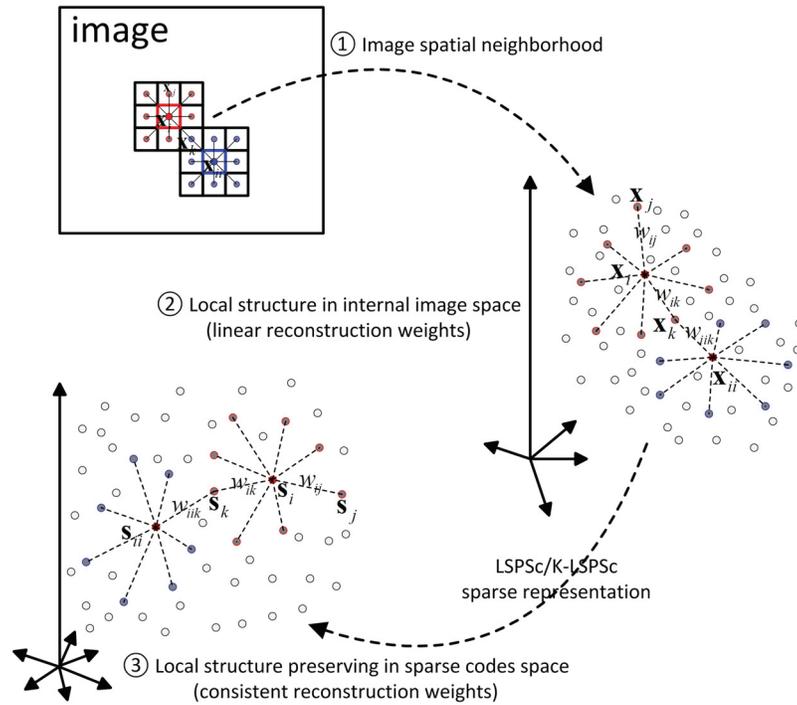
The L1 penalty is used to formulate the sparse coding, instead of solving the NP-hard problem of minimized L0 regularization. The first term in Eq 1 is the reconstruction error. The second term is used to control the sparsity of the sparse codes  $\mathbf{S}$  and the reconstruction error through the proportion parameter  $\lambda$ . There is a norm constraint for basis to some constant  $c$ , which retains the variation of coefficients for each basis at the same level. The dictionary basis  $\mathbf{B}$  and the sparse codes  $\mathbf{S}$  should be optimized simultaneously [20,40].

### Local structure preserving sparse coding

The dominated sparse pursuit results in the loss of the spatial local manifold in sparse codes of the patches. To preserve such spatial locality, it is necessary to incorporate an effective prior (denoted as a regularization term) into the reconstruction process [41–43]. To maintain the structural relationship of local neighborhood in the internal image space with the corresponding sparse codes space, we explore the local manifold among patches as a prior, and introduce the neighborhood-preserved embeddings to the objective of sparse coding. As shown in Fig 2.

In the existing manifold learning methods, local linear embedding assumes that each data point and its neighbors conform to or close to a locally linear manifold [44–47], and recovers global nonlinear structure from locally linear fits [48,49]. By exploiting the local symmetries of linear reconstructions, its optimizations do not involve local minima. Inspired by these, we characterize the local geometry of the patches in image by linear coefficients that reconstruct each patch from its neighbors.

Denote the local manifold matrix corresponding to these patches as  $\mathbf{W} \in R^{N \times N}$ , whose entry  $w_{ij}$  measures the reconstruction weights from a surrounding patch to its center patch. The weights are evaluated by minimizing the local structure reconstruction error. With the



**Fig 2. The schematic diagram of LSPSc.**

<https://doi.org/10.1371/journal.pone.0173613.g002>

regularization prior, we can formulate the LSPSc as:

$$\min_{\mathbf{B}, \mathbf{S}} \sum_i \|\mathbf{x}_i - \mathbf{B}\mathbf{s}_i\|^2 + \lambda \sum_i \|\mathbf{s}_i\|_1 + \beta \sum_i \|\mathbf{s}_i - \sum_j w_{ij} \mathbf{s}_j\|^2 \quad (2)$$

subject to  $\|\mathbf{b}_i\|^2 \leq c, i = 1, 2, \dots, M, c$  is the constraint constant and is set to 1 in this paper. Here is  $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_N]^T$  and  $\mathbf{w}_i = [w_{i1}, w_{i2}, \dots, w_{iN}]$ , which subject to the constraints  $\sum_{\mathbf{x}_j \in \Omega_i} w_{ij} = 1$  and  $w_{ij} = 0$  for any  $\mathbf{x}_j \notin \Omega_i$ . For each patch  $\mathbf{x}_i$ , the set  $\Omega_i$  is its spatial  $K$  nearest neighborhood. Defining the constraint weights matrix of the local manifold  $\mathbf{M} = (\mathbf{I} - \mathbf{W})^T (\mathbf{I} - \mathbf{W})$ , we can rewrite Eq 2 into matrix form

$$\min_{\mathbf{B}, \mathbf{S}} \|\mathbf{X} - \mathbf{B}\mathbf{S}\|^2 + \lambda \sum_i \|\mathbf{s}_i\|_1 + \beta \text{tr}(\mathbf{S}\mathbf{M}\mathbf{S}^T) \quad (3)$$

### Kernel local structure preserving sparse coding

Motivated by the nonlinear generalization performance of kernel methods, we extend LSPSc to a kernel version (K-LSPSc). An implicit mapping function  $\phi(\cdot)$  maps the input samples and basis to the high dimensional kernel space. K-LSPSc seeks the sparse codes for the mapped sample under the mapped basis.

$$\begin{aligned} \mathbf{X} \overrightarrow{\mathbf{X}}_\phi &= [\phi(\mathbf{x}_1), \phi(\mathbf{x}_2), \dots, \phi(\mathbf{x}_N)] \\ \mathbf{B} \overrightarrow{\mathbf{B}}_\phi &= [\phi(\mathbf{b}_1), \phi(\mathbf{b}_2), \dots, \phi(\mathbf{b}_M)] \end{aligned} \quad (4)$$

The local manifold reconstruction weights in LSPSc should also be achieved in kernel space. In other words, we need to obtain the geometry relationships of neighboring patches in kernel space by computing the local manifold reconstruction weights among samples in  $\mathbf{X}_\phi$ . The

spatial neighborhood is supposed to map unchanged [50].

$$\mathbf{W}\overrightarrow{\mathbf{W}}_{\phi} \tag{5}$$

With the local structure prior, we essentially pursue the sparse coefficients  $\hat{\mathbf{S}}$  of LSPSc in kernel space. The formulation of K-LSPSc is expressed as in Eq 6, subject to  $\|\phi(\mathbf{b}_i)\|^2 \leq c, i = 1, 2, \dots, M$ . We also constraint  $\sum_{\mathbf{x}_j \in \Omega_i} w_{\phi ij} = 1$  in kernel space.

$$\min_{\mathbf{B}_{\phi}, \hat{\mathbf{S}}} \|\mathbf{X}_{\phi} - \mathbf{B}_{\phi} \hat{\mathbf{S}}\|^2 + \lambda \sum_i \|\hat{\mathbf{s}}_i\|_1 + \beta \sum_i \|\hat{\mathbf{s}}_i - \sum_j w_{\phi ij} \hat{\mathbf{s}}_j\|^2 \tag{6}$$

**Relationship between LSPSc and K-LSPSc.** In kernel space, samples in different categories are linearly separable, which enhances the discriminative ability of the sparse coding and boosts target recognition performance. On the other hand, LSPSc assumes that the local structure of the patches in natural image satisfies a linear relationship, but there are more nonlinear structures in practical image data. K-LSPSc supposes local patches in image to meet the linear reconstruction in high dimensional kernel space. Thus, when the data in kernel space are non-linearly mapped to a low dimensional space, the data structure is complex, which is according with the distribution of actual data. Therefore, by the kernel skill the local reconstruction error may be reduced, and we can get the structure sparse quantization for the signals more accurate and discriminative.

### The implementation of LSPSc and K-LSPSc

The optimization problems in Eq 2 and Eq 6 can be solved in two core aspects: the calculation of structure matrix ( $\mathbf{W}$  and  $\mathbf{W}_{\phi}$ ) and optimization scheme of sparse coefficient and dictionary ( $\mathbf{S}, \mathbf{B}$  and  $\hat{\mathbf{S}}, \mathbf{B}_{\phi}$ ).

**Calculation of  $\mathbf{W}$  and  $\mathbf{W}_{\phi}$ .** The local manifold reconstruction weights  $\mathbf{W}$  from the neighbors  $\{\mathbf{x}_j | j = 1, 2, \dots, K\}$  to a center patch  $\mathbf{x}_i$  are computed by minimizing the reconstruction error, which is measured by the cost function:

$$\min_{\mathbf{W}} \sum_{i=1}^N \|\mathbf{x}_i - \sum_{j=1}^K w_{ij} \mathbf{x}_j\|^2, \quad s.t. \sum_j w_{ij} = 1 \tag{7}$$

Here is  $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_N]$  and  $\mathbf{w}_i = [w_{i1}, w_{i2}, \dots, w_{iK}]^T$ . Minimizing the cost function subject to the constraint  $\mathbf{e}^T \mathbf{w}_i = 1$  is a constrained least squares problem [47–49].

Setting  $f(\mathbf{w}_i) = \|\mathbf{x}_i - \sum_{j=1}^K w_{ij} \mathbf{x}_j\|^2$  and  $\mathbf{G}_i = [\mathbf{x}_i - \mathbf{x}_{i1}, \dots, \mathbf{x}_i - \mathbf{x}_{iK}]$ , there is  $f(\mathbf{w}_i) = \|\mathbf{G}_i \mathbf{w}_i\|^2 = \mathbf{w}_i^T \mathbf{G}_i^T \mathbf{G}_i \mathbf{w}_i$ . So the optimization problem in Eq 7 can be converted as  $\min \mathbf{w}_i^T \mathbf{G}_i^T \mathbf{G}_i \mathbf{w}_i$  subject to  $\mathbf{e}^T \mathbf{w}_i = 1$ . Consider the Lagrangian  $\ell(\mathbf{w}_i) = \mathbf{w}_i^T \mathbf{G}_i^T \mathbf{G}_i \mathbf{w}_i + \eta(\mathbf{w}_i^T \mathbf{e} - 1)$ . The optimal weights are gained by requiring the first derivative of  $\ell(\mathbf{w}_i)$  versus  $\mathbf{w}_i$  to be zero:  $\partial \ell(\mathbf{w}_i) / \partial \mathbf{w}_i = \mathbf{G}_i^T \mathbf{G}_i \mathbf{w}_i + \eta \mathbf{e} = 0$ . Then solve Eq 8 and rescale the weights to get the desired solution  $\mathbf{W}$ . Here is  $\mathbf{G}_i^T \mathbf{G}_i = [(\mathbf{x}_i - \mathbf{x}_{it1})(\mathbf{x}_i - \mathbf{x}_{it2})]_{t1, t2=1, \dots, K}$ .

$$\mathbf{G}_i^T \mathbf{G}_i \mathbf{w}_i = \mathbf{e} \tag{8}$$

In kernel space, each sample  $\mathbf{x}_i$  is transformed to  $\phi(\mathbf{x}_i)$ . The reconstruction error can then be minimized as follows:

$$\min_{\mathbf{W}_{\phi}} \sum_{i=1}^N \|\phi(\mathbf{x}_i) - \sum_{j=1}^K w_{\phi ij} \phi(\mathbf{x}_j)\|^2, \quad s.t. \sum_j w_{\phi ij} = 1 \tag{9}$$

Setting  $\mathbf{w}_{\phi i} = [w_{\phi i1}, w_{\phi i2}, \dots, w_{\phi iK}]^T$ ,  $\mathbf{G}_{\phi i} = [\phi(\mathbf{x}_i) - \phi(\mathbf{x}_{i1}), \dots, \phi(\mathbf{x}_i) - \phi(\mathbf{x}_{iK})]$ . In accordance with

the above principle, the desired solution in kernel space  $\mathbf{W}_\phi$  can be obtained by solving  $\mathbf{G}_{\phi i}^T \mathbf{G}_{\phi i} \mathbf{w}_{\phi i} = \mathbf{e}$  and rescaling the weights. In kernel trick, the kernel function is defined as the inner product of two mapping functions  $k(x, y) = \langle \phi(x), \phi(y) \rangle$ , which should be positive definite. So there is  $\mathbf{G}_{\phi i}^T \mathbf{G}_{\phi i} = [1 - k(\mathbf{x}_i, \mathbf{x}_{i_1}) - k(\mathbf{x}_i, \mathbf{x}_{i_2}) + k(\mathbf{x}_{i_1}, \mathbf{x}_{i_2})]_{i_1, i_2=1, \dots, K}$ .

**Implementation of LSPSc and K-LSPSc.** DR solves the optimization based on proximal point algorithm and accelerated proximal gradient line search method, which has high computational cost. We employ the efficient strategy of feature-sign search algorithm in [39] to solve the optimization objective of Eq 3 by alternately optimizing  $\mathbf{B}$  and  $\mathbf{S}$  while holding the other fixed. K-LSPSc is the same as LSPSc except for the kernel mapping. Thus, we can use the same method to optimize the objective of K-LSPSc.

First, learn the coefficients  $\mathbf{S}$  and  $\hat{\mathbf{S}}$ . In LSPSc,  $\mathbf{B}$  is fixed and the optimization problem of Eq 2 is equivalent to a regularized least squares problem. We optimize each  $\mathbf{s}_i$  individually with all the remaining sparse codes  $\mathbf{s}_j (j \neq i)$  fixed in Eq 10. By replacing the sparse reconstruction error term in [39] with  $J(\mathbf{s}_i)$ , the efficient sparse coding algorithm can be easily extended to optimize LSPSc coefficients in Eq 10.

$$\min_{\mathbf{s}_i} J(\mathbf{s}_i) + \lambda \|\mathbf{s}_i\|_1 \tag{10}$$

$$J(\mathbf{s}_i) = \|\mathbf{x}_i - \mathbf{B}\mathbf{s}_i\|^2 + \beta \|\mathbf{s}_i - \sum_j w_{ij} \mathbf{s}_j\|^2$$

Due to the constraint  $\sum_j w_{ij} = 1$ , there is  $\mathbf{s}_i^T \mathbf{s}_i = \sum_j w_{ij} \mathbf{s}_i^T \mathbf{s}_j$ , so  $J(\mathbf{s}_i)$  can be transformed into Eq 11.

$$J(\mathbf{s}_i) = \mathbf{s}_i^T \mathbf{B}^T \mathbf{B} \mathbf{s}_i - \mathbf{x}_i^T \mathbf{B} \mathbf{s}_i - \mathbf{s}_i^T \mathbf{B}^T \mathbf{x}_i + \beta [\mathbf{s}_i^T (\mathbf{S} \mathbf{L}_i) + (\mathbf{S} \mathbf{L}_i)^T \mathbf{s}_i - \mathbf{s}_i^T \mathbf{L}_{ii} \mathbf{s}_i] + f(\mathbf{s}_i) \tag{11}$$

Here,  $f(\mathbf{s}_i) = \mathbf{x}_i^T \mathbf{x}_i + \beta (\sum_j w_{ij} \mathbf{s}_j)^T (\sum_j w_{ij} \mathbf{s}_j)$  and  $\mathbf{L} = \mathbf{I} - \mathbf{W}$ , where  $\mathbf{I}$  is an identity matrix.  $\mathbf{L}_i$  is the  $i^{th}$  column of matrix  $\mathbf{L}$ . Actually,  $\mathbf{L}_{ii} = 1$  due to  $w_{ii} = 0$ . The details of solving Eq 10 are listed in Algorithm 1, in which the first and second derivative of  $J(\mathbf{s}_i)$  versus  $\mathbf{s}_i$  can be expressed as Eq 12.

$$\frac{\partial J(\mathbf{s}_i)}{\partial \mathbf{s}_i} = -2\mathbf{B}^T \mathbf{x}_i + 2\mathbf{B}^T \mathbf{B} \mathbf{s}_i + \beta (2\mathbf{S}_{-i} \mathbf{L}_{i,-i} - 2\mathbf{L}_{ii} \mathbf{s}_i) \tag{12}$$

$$\frac{\partial^2 J(\mathbf{s}_i)}{\partial \mathbf{s}_i^2} = 2\mathbf{B}^T \mathbf{B} - 2\beta \mathbf{L}_{ii} \mathbf{I}$$

Here,  $\mathbf{S}_{-i}$  is the submatrix after removing the  $i^{th}$  column from  $\mathbf{S}$ , and  $\mathbf{L}_{i,-i}$  is the subvector after removing the  $i^{th}$  entry from  $\mathbf{L}_i$ . Some recent works show that the coefficients  $\mathbf{S}$  initialized with some heuristics can achieve faster convergence [20]. Thus,  $\mathbf{S}$  is initialized with the results of general sparse coding in Algorithm 1.

In K-LSPSc, Eq 10 should be implemented in kernel space, which can be expressed as follows:  $\mathbf{S} \hat{\mathbf{S}}, J(\mathbf{s}_i) \vec{J}_\phi(\hat{\mathbf{s}}_i)$ .

$$J_\phi(\hat{\mathbf{s}}_i) = \|\phi(\mathbf{x}_i) - \mathbf{B}_\phi \hat{\mathbf{s}}_i\|^2 + \beta \|\hat{\mathbf{s}}_i - \sum_j w_{\phi ij} \hat{\mathbf{s}}_j\|^2. \tag{13}$$

There is also the constraint  $\sum_j w_{\phi ij} = 1$  in kernel space, so  $J_\phi(\hat{\mathbf{s}}_i)$  can be further transformed to Eq 14.

$$J_\phi(\hat{\mathbf{s}}_i) = \hat{\mathbf{s}}_i^T \mathbf{B}_\phi^T \mathbf{B}_\phi \hat{\mathbf{s}}_i - \phi(\mathbf{x}_i)^T \mathbf{B}_\phi \hat{\mathbf{s}}_i - \hat{\mathbf{s}}_i^T \mathbf{B}_\phi^T \phi(\mathbf{x}_i) + \beta [(\hat{\mathbf{s}}_i^T (\hat{\mathbf{S}} \mathbf{L}_{\phi i}) + (\hat{\mathbf{S}} \mathbf{L}_{\phi i})^T \hat{\mathbf{s}}_i - \hat{\mathbf{s}}_i^T \mathbf{L}_{\phi ii} \hat{\mathbf{s}}_i)] + f(\hat{\mathbf{s}}_i). \tag{14}$$

Here,  $f(\hat{\mathbf{s}}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_i) + \beta(\sum_j w_{\phi ij} \hat{\mathbf{s}}_j)^T (\sum_j w_{\phi ij} \hat{\mathbf{s}}_j)$  and  $\mathbf{L}_\phi = \mathbf{I} - \mathbf{W}_\phi$ . With the kernel trick, we have  $\mathbf{B}_\phi^T \mathbf{B}_\phi = \{\phi(\mathbf{b}_m)^T \phi(\mathbf{b}_n)\}_{m,n=1,2,\dots,M}$ ,  $\phi(\mathbf{b}_m)^T \phi(\mathbf{b}_n) = \langle \mathbf{b}_m, \mathbf{b}_n \rangle = k(\mathbf{b}_m, \mathbf{b}_n)$  and  $\mathbf{B}_\phi^T \phi(\mathbf{x}_i) = [\phi(\mathbf{b}_1)^T \phi(\mathbf{x}_i), \dots, \phi(\mathbf{b}_m)^T \phi(\mathbf{x}_i), \dots, \phi(\mathbf{b}_M)^T \phi(\mathbf{x}_i)]_{m=1,2,\dots,M}$ ,  $\phi(\mathbf{b}_m)^T \phi(\mathbf{x}_i) = \langle \mathbf{b}_m, \mathbf{x}_i \rangle = k(\mathbf{b}_m, \mathbf{x}_i)$ . All these subitems are calculable, so the efficient sparse coding algorithm can also be employed to get KLSPSc coefficients  $\hat{\mathbf{S}}$ .

Second, learn the dictionary basis  $\mathbf{B}$  and  $\mathbf{B}_\phi$ . In LSPSc,  $\mathbf{S}$  is fixed and the optimization problem of Eq 3 can be rewritten as Eq 15, which is a least squares problem with quadratic constraints. Following [39], by solving a Lagrange dual through conjugate gradient method, we optimize the basis  $\mathbf{B}$  as Eq 16.

$$\min_{\mathbf{B}} \|\mathbf{X} - \mathbf{BS}\|^2, \quad s.t. \|\mathbf{b}_i\|^2 \leq c, \quad i = 1, 2, \dots, M \tag{15}$$

$$\mathbf{B}^T = (\mathbf{SS}^T + \mathbf{\Lambda})^{-1} (\mathbf{XS}^T)^T \tag{16}$$

Here,  $\mathbf{\Lambda}$  is a diagonal matrix of the dual variable, which is set as an identity matrix in this paper.

In K-LSPSc, the optimization problem of Eq 6 can be solved with the same method, and the basis  $\mathbf{B}_\phi$  is optimized as:

$$\mathbf{B}_\phi^T = (\hat{\mathbf{S}}\hat{\mathbf{S}}^T + \mathbf{\Lambda})^{-1} (\phi(\mathbf{X})\hat{\mathbf{S}}^T)^T \tag{17}$$

It can be seen that the updated basis in kernel space  $\mathbf{B}_\phi$  cannot be obtained numerically for the implicit mapping function in Eq 17. Actually, in the  $n^{th}$   $\hat{\mathbf{S}}$  updating iteration by the efficient sparse coding algorithm, we can use the result of  $\mathbf{B}_\phi$  in the  $(n-1)^{th}$  updating iteration to compute the part in Eq 14:  $\mathbf{B}_\phi^T \phi(\mathbf{x}_i) = [(\hat{\mathbf{S}}\hat{\mathbf{S}}^T + \mathbf{\Lambda})^{-1} (\phi(\mathbf{X})\hat{\mathbf{S}}^T)^T]_{last} \phi(\mathbf{x}_i) = [(\hat{\mathbf{S}}\hat{\mathbf{S}}^T + \mathbf{\Lambda})^{-1} \hat{\mathbf{S}}]_{last} \phi(\mathbf{X})^T \phi(\mathbf{x}_i) = [(\hat{\mathbf{S}}\hat{\mathbf{S}}^T + \mathbf{\Lambda})^{-1} \hat{\mathbf{S}}]_{last} [k(\mathbf{x}_1, \mathbf{x}_i), \dots, k(\mathbf{x}_i, \mathbf{x}_i), \dots, k(\mathbf{x}_N, \mathbf{x}_i)]$  and  $\mathbf{B}_\phi^T \mathbf{B}_\phi = [(\hat{\mathbf{S}}\hat{\mathbf{S}}^T + \mathbf{\Lambda})^{-1} (\phi(\mathbf{X})\hat{\mathbf{S}}^T)^T]_{last} [(\hat{\mathbf{S}}\hat{\mathbf{S}}^T + \mathbf{\Lambda})^{-1} \hat{\mathbf{S}}]_{last} \phi(\mathbf{X})^T \phi(\mathbf{X}) [(\hat{\mathbf{S}}\hat{\mathbf{S}}^T + \mathbf{\Lambda})^{-1} \hat{\mathbf{S}}]_{last}^T$ .

Different with the linear kernel used in DR, we choose the RBF kernel  $k(x, y) = \exp(-\|x - y\|^2 / \sigma_k^2)$ , which can estimate the nonlinear similarity between two signals. The Gaussian variance  $\sigma_k$  is a constant and is set as 1.

**Algorithm 1** Pseudocode for Feature-Sign Search Algorithm in Solving Eq 10

1: **Input:**  $\mathbf{x}_i$  the  $i^{th}$  sample in  $\mathbf{X}$ ; the dictionary  $\mathbf{B}$ ; initial sparse coding  $\mathbf{S}$  and  $\lambda, \beta$  of LSPSc.

**Output:** The LSPSc sparse codes  $\mathbf{s}_i$  of sample  $\mathbf{x}_i$ .

2: **Initialize:**  $\mathbf{S}$  is initialized as the classical sparse coding, sign set  $\boldsymbol{\theta} = \text{sign}(\mathbf{s}_i)$ , active set  $\Omega = \text{find}(\mathbf{s}_i \neq 0)$ .

3: Calculate the local manifold reconstruction weights  $\mathbf{w}_i$  by Eq 8 to get  $\mathbf{L}_i$ .

4: From zero coefficients of  $\mathbf{s}_i$ , select  $rm = \arg \max_r |\Delta_{s_i}^r|$ ,  $\Delta_{s_i} = \partial J(\mathbf{s}_i) / \partial \mathbf{s}_i$ . Denote  $\mathbf{v}^r$  as the  $r^{th}$  entries of vector  $\mathbf{v}$ .

If  $\Delta_{s_i}^{rm} > \lambda$ , set  $\boldsymbol{\theta}^{rm} = -1$ , active set  $\Omega = \{rm\} \cup \Omega$ .

If  $\Delta_{s_i}^{rm} < -\lambda$ , set  $\boldsymbol{\theta}^{rm} = 1$ , active set  $\Omega = \{rm\} \cup \Omega$ .

5: **Feature-sign step:**

Compute the analytical solution to the resulting unconstrained QP:

$$\min_{\mathbf{s}_i, \Omega} J(\mathbf{s}_i, \Omega) + \lambda \boldsymbol{\theta}_\Omega^T \mathbf{s}_i, \Omega.$$

$$\mathbf{s}_i, \Omega_{new} = ((\Delta_{\mathbf{s}_i, \Omega})^{-1} (2\mathbf{B}_\Omega^T \mathbf{x}_i - 2\beta(\mathbf{S}_{-i} \mathbf{L}_{i,-i})_\Omega - \lambda \boldsymbol{\theta}_\Omega))$$

Denote  $\mathbf{v}_\Omega$  as the subvector of vector  $\mathbf{v}$  corresponding to the active set  $\Omega$ , and  $\mathbf{M}_\Omega$  as the submatrix of  $\mathbf{M}$  that contains only the columns corresponding to the active set  $\Omega$ .  $\Delta_{\mathbf{s}_i, \Omega} = \partial^2 J(\mathbf{s}_i) / \partial \mathbf{s}_i^2$ .

Perform a discrete line search on the closed line segment from  $\mathbf{s}_i, \Omega$  to

$\mathbf{s}_i, \Omega_{new}$ :

Check the objective value at  $\mathbf{s}_{i\Omega_{new}}$  and all points where any coefficient changes sign.

Update  $\mathbf{s}_{i\Omega}$  (and the corresponding entries in  $\mathbf{s}_{i\Omega}$ ) to the point with the lowest objective value.

Remove zero coefficients of  $\mathbf{s}_{i\Omega}$  from the active set  $\Omega$  and update  $\theta = \text{sign}(\mathbf{s}_i)$ .

**6: Check the optimality conditions:**

(a) Optimality condition for nonzero coefficients:  $\Delta_{s_i^r} + \lambda \text{sign}(s_i^r) = 0$ ,  $\forall s_i^r \neq 0$ .

If condition (a) is not satisfied, go to Step 5 (without any new activation); else check condition (b).

(b) Optimality condition for zero coefficients:  $|\Delta_{s_i^r}| \leq \lambda$ ,  $\forall s_i^r = 0$ .

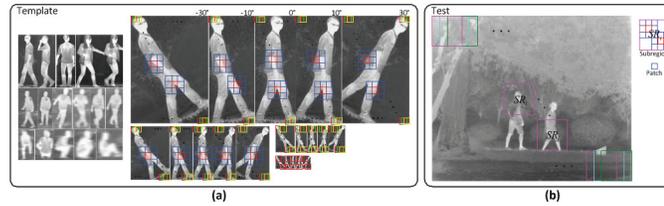
If condition (b) is not satisfied, go to Step 4; otherwise return  $\mathbf{s}_i$  as the solution, and update the  $\mathbf{S}$  of LSPSc with current  $\mathbf{s}_i$ .

In Algorithm 1, each step reduces the objective

$$f(\mathbf{s}_i) = J(\mathbf{s}_i) + \lambda \|\mathbf{s}_i\|_1, J(\mathbf{s}_i) = \|\mathbf{x}_i - \mathbf{B}\mathbf{s}_i\|^2 + \beta \|\mathbf{s}_i - \sum_j w_{ij} \mathbf{s}_j\|^2$$

and that the overall algorithm always converges to the optimal solution Eq 10. Consider optimization problem Eq 10 augmented with the additional constraint that  $\mathbf{s}_i$  is consistent with a given active set and sign vector. If the current coefficients  $\mathbf{s}_{ic}$  are consistent with the active set and sign vector, but are not optimal for the augmented problem at the start of Step 5, the feature-sign step is guaranteed to strictly reduce the objective; If the coefficients  $\mathbf{s}_{ic}$  at the start of Step 4 are optimal for the augmented problem, but are not optimal for problem Eq 10, the feature-sign step is guaranteed to strictly reduce the objective. The algorithm converges to a global optimum of the optimization problem Eq 10 in a finite number of steps.

- Let  $\hat{\mathbf{s}}_{ic}$  be the subvector of  $\mathbf{s}_{ic}$  corresponding to coefficients in the given active set  $\Omega$ . In Step 5, consider a smooth quadratic function  $\tilde{f}(\hat{\mathbf{s}}_i) = \|\mathbf{x}_i - \hat{\mathbf{B}}\hat{\mathbf{s}}_i\|^2 + \beta \|\hat{\mathbf{s}}_i - \sum_{ij} w_{ij} \hat{\mathbf{s}}_j\|^2 + \lambda \hat{\theta}^T \hat{\mathbf{s}}_i$ . Since  $\hat{\mathbf{s}}_{ic}$  is not an optimal point of  $\tilde{f}$ , we have  $\tilde{f}(\hat{\mathbf{s}}_{i\text{new}}) < \tilde{f}(\hat{\mathbf{s}}_{ic})$ . Now consider the two possible cases: (1) if  $\hat{\mathbf{s}}_{i\text{new}}$  is consistent with the given active set and sign vector, updating  $\hat{\mathbf{s}}_i := \hat{\mathbf{s}}_{i\text{new}}$  strictly decreases the objective; (2) if  $\hat{\mathbf{s}}_{i\text{new}}$  is not consistent with the given active set and sign vector, let  $\hat{\mathbf{s}}_{id}$  be the first zero-crossing point (where any coefficient changes its sign) on a line segment from  $\hat{\mathbf{s}}_{ic}$  to  $\hat{\mathbf{s}}_{i\text{new}}$ , then clearly  $\hat{\mathbf{s}}_{ic} \neq \hat{\mathbf{s}}_{id}$ , and  $\tilde{f}(\hat{\mathbf{s}}_{id}) < \tilde{f}(\hat{\mathbf{s}}_{ic})$  by convexity of  $\tilde{f}$ , thus we finally have  $f(\hat{\mathbf{s}}_{id}) = \tilde{f}(\hat{\mathbf{s}}_{id}) < \tilde{f}(\hat{\mathbf{s}}_{ic}) = f(\hat{\mathbf{s}}_{ic})$ . Therefore, the discrete line search described in Step 5 ensures a decrease in the objective value.
- Since  $\mathbf{s}_{ic}$  is optimal for the augmented problem, it satisfies optimality condition (a), but not (b); thus, in Step 4, there is some  $r$ , such that  $|\Delta_{s_i^r}| > \lambda$ ; this  $r$ -th coefficient is activated, and  $r$  is added to the active set  $\Omega$ . In Step 5, consider the smooth quadratic function  $\tilde{f}(\hat{\mathbf{s}}_i) = \|\mathbf{x}_i - \hat{\mathbf{B}}\hat{\mathbf{s}}_i\|^2 + \beta \|\hat{\mathbf{s}}_i - \sum_{ij} w_{ij} \hat{\mathbf{s}}_j\|^2 + \lambda \hat{\theta}^T \hat{\mathbf{s}}_i$ . Observe that (1) since a Taylor expansion of  $\tilde{f}$  around  $\hat{\mathbf{s}}_i = \hat{\mathbf{s}}_{ic}$  has a first order term in  $\mathbf{s}_i^r$  only (using condition 6(a) for the other coefficients), we have that any direction that locally decreases  $\tilde{f}(\hat{\mathbf{s}}_i)$  must be consistent with the sign of the activated  $\mathbf{s}_i^r$ , and, (2) since  $\hat{\mathbf{s}}_{ic}$  is not an optimal point of  $\tilde{f}(\hat{\mathbf{s}}_i)$ ,  $\tilde{f}(\hat{\mathbf{s}}_i)$  must decrease locally near  $\hat{\mathbf{s}}_i = \hat{\mathbf{s}}_{ic}$  along the direction from  $\hat{\mathbf{s}}_{ic}$  to  $\hat{\mathbf{s}}_{i\text{new}}$ . From Eq 1 and Eq 2, the line search direction  $\hat{\mathbf{s}}_{ic}$  to  $\hat{\mathbf{s}}_{i\text{new}}$  must be consistent with the sign of the activated  $\mathbf{s}_i^r$ . Finally, since  $\tilde{f}(\hat{\mathbf{s}}_i) = f(\hat{\mathbf{s}}_i)$  when  $\hat{\mathbf{s}}_i$  is consistent with the active set  $\Omega$ , either  $\hat{\mathbf{s}}_{i\text{new}}$  is consistent, or the first zero-crossing from  $\hat{\mathbf{s}}_{ic}$  to  $\hat{\mathbf{s}}_{i\text{new}}$  has a lower objective value (similar argument to A).
- From the above arguments, it follows that the steps always strictly reduce the objective  $f(\mathbf{s}_i)$ . At the start of Step 4,  $\mathbf{s}_i$  either satisfies optimality condition 6(a) or is  $\vec{0}$ ; in either case,  $\mathbf{s}_i$  is



**Fig 3. Diagram of the LSSM for human recognition.** (a) Sampling from the expanded template set for learning the local sparse structure dictionary; (b) LSSM target recognition based on LSPSc.

<https://doi.org/10.1371/journal.pone.0173613.g003>

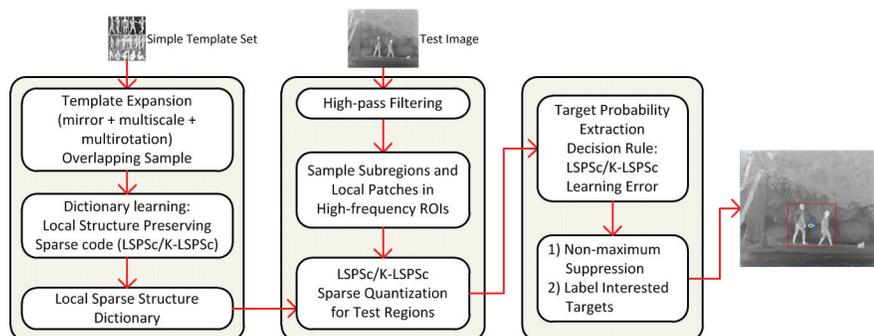
consistent with the current active set and sign vector, and must be optimal for the augmented problem described in the above arguments. Since the number of all possible active sets and coefficient signs is finite, and since no pair can be repeated (because the objective value is strictly decreasing), the outer loop of Steps 4-6(b) cannot repeat indefinitely. Now, it suffices to show that a finite number of steps is needed to reach Step 6(b) from Step 4. This is true because the inner loop of Steps 5-6(a) always results in either an exit to Step 6 (b) or a decrease in the size of the active set.

### Local sparse structure matching model

The local sparse structure matching (LSSM) model based on LSPSc and K-LSPSc is designed for robust recognition of general infrared targets. Fig 3 illustrates this idea using human recognition as an example. Fig 4 shows an overview of our proposed framework, which mainly includes dictionary generation, sparse quantization for test regions, and target probability extraction.

**Local sparse structure dictionary generation.** Expand the simple template set by scaling and rotating every image and its mirror in the template set (Fig 3(A)), we can learn the dictionary of interested targets with multi-scale and multi-rotation information. The scale is from 0.2 to 2 by steps of 0.2; the rotation is from  $-90^\circ$  to  $90^\circ$  by steps of  $10^\circ$ . Obviously it is time consuming, but the dictionary is pre-learned, which is before sparse quantization of test images. It has no effect on the efficiency of target identification, but improves the robustness of the dictionary.

Sample the patches (small windows in Fig 3(A)) by pixels (with step of one pixel) from the expanded template images. Fix each patch size  $p_s$  at 12 and consider its nonoverlapping neighbors. Calculate the neighborhood reconstruction weights among each patch and its eight neighboring patches by Eq 8. We iteratively optimize the sparse coefficients (by Algorithm 1) and dictionary (by Eq 16) while holding the other fixed.



**Fig 4. Framework of the LSSM model.**

<https://doi.org/10.1371/journal.pone.0173613.g004>

**LSPSc/K-LSPSc sparse quantization and target probability extraction.** Take a similar approach to sample subregions (large windows in Fig 3(B)) from a test image, which is mainly used to study the local sparsity and structure information. The subregion size  $r_s$  is set as 60 by considering both sparsity and locality. Each subregion is divided into a set of nonoverlapping patches (same size as template patches). Using the local sparse structure dictionary of the template, the sparse representation of all patches in subregion is quantized according to LSPSc/K-LSPSc. Through the K-LSPSc dictionary  $\mathbf{B}_\phi = [\phi(\mathbf{X})\hat{\mathbf{S}}^T(\hat{\mathbf{S}}\hat{\mathbf{S}}^T + \mathbf{\Lambda})^{-1}]_{template}$  is nonnumeric, the expression  $\mathbf{B}_\phi^T\phi(\mathbf{x}_i)$  and  $\mathbf{B}_\phi^T\mathbf{B}_\phi$  in  $\Delta_{\hat{s}_i}$  and  $\Delta_{\hat{s}_i, \hat{s}_i}$  in Algorithm 1 can be formulated as  $\mathbf{B}_\phi^T\phi(\mathbf{x}_i) = [(\hat{\mathbf{S}}\hat{\mathbf{S}}^T + \mathbf{\Lambda})^{-1}\hat{\mathbf{S}}]_{template}k(\mathbf{X}_{template}, \mathbf{x}_i)$  and  $\mathbf{B}_\phi^T\mathbf{B}_\phi = [(\hat{\mathbf{S}}\hat{\mathbf{S}}^T + \mathbf{\Lambda})^{-1}\hat{\mathbf{S}}]_{template}k(\mathbf{X}_{template}, \mathbf{X}_{template})[\hat{\mathbf{S}}^T(\hat{\mathbf{S}}\hat{\mathbf{S}}^T + \mathbf{\Lambda})^{-1}]_{template}$ , they are calculable and numerical. Here,  $\mathbf{x}_i$  is a sample from test patches.

We present a new definition of target probability for each subregion, which is used to determine the matching degree between test local region and template. The target probability  $\rho_t$  is defined by the average LSPSc/K-LSPSc objective value  $H_t$  of all patches in each subregion  $SR_t$ , as in Eq 17.  $H_t$  is also called ‘‘matching error’’ hereinafter.

$$H_t = \frac{1}{N} \sum_{SR_t} \{ \min_s \|\mathbf{X} - \mathbf{B}\mathbf{S}\|^2 + \lambda \sum_i \|s_i\|_1 + \beta tr(\mathbf{S}\mathbf{M}\mathbf{S}^T) \} \tag{18}$$

$$\rho_t = \exp(-H_t^2/2\sigma^2)$$

Here  $\mathbf{X}$  is a sample set of all patches in a test subregion. The target probability in the kernel skill is formulated as:

$$\hat{H}_t = \frac{1}{N} \sum_{SR_t} \{ \min_s \|\mathbf{X}_\phi - \mathbf{B}_\phi\hat{\mathbf{S}}\|^2 + \lambda \sum_i \|\hat{s}_i\|_1 + \beta tr(\hat{\mathbf{S}}\mathbf{M}_\phi\hat{\mathbf{S}}^T) \} \tag{19}$$

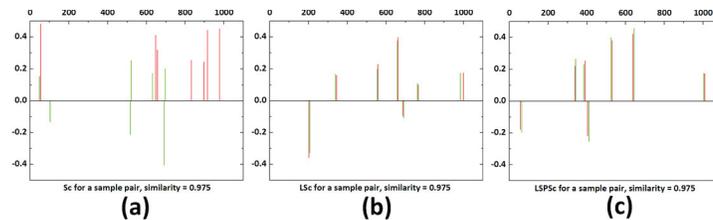
$$\hat{\rho}_t = \exp(-\hat{H}_t^2/2\hat{\sigma}^2)$$

The Gaussian variables  $\sigma$  and  $\hat{\sigma}$  control the degree of attenuation from target probability to matching error value. They are set as the average LSPSc/K-LSPSc matching error according to target category, and are regarded as priori constants in LSSM. After the target probability tests are performed, we employ the approaches of significance tests and non-maxima suppression [51] for final detection.

The local sparse structure dictionary contains sparse elements and spatial local manifolds of template. From Eqs 18 and 19, the second term constrains the consistency between test patches and template basis; the third term preserves the consistent neighborhood reconstruction relationship between test patches and their sparse codes. The sparse coefficient  $\mathbf{S}$  and  $\hat{\mathbf{S}}$  rely on the template dictionary, while the constraint weights of local structure  $\mathbf{M}$  and  $\mathbf{M}_\phi$  rely on test regions, so the minimum of the third term ensures a similar local sparse structure between test regions and template set. Therefore, only the test regions, which have patches matching with template dictionary basis and local sparse structure similar with template subregions both, deserve low sparse quantization error (first term), few dictionary basis response (second term), and low constrained local sparse structure error (third term). Then these regions get large target probability.

### Analysis of LSPSc/K-LSPSc in LSSM

**Analysis of the ability of similarity preserving by LSPSc.** As a local sparse matching approach, our model requires the stability of sparse representation. It should suppress the noise interference and constrain the consistency of sparse representations among similar



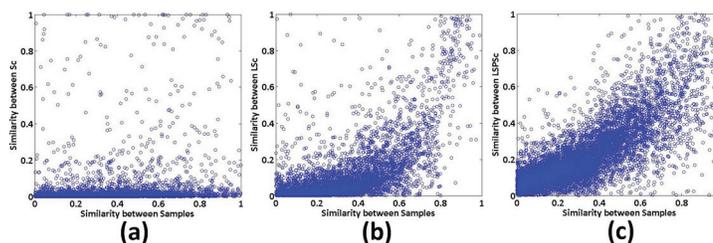
**Fig 5. Comparison of the sparse codes of similar samples (a pair of neighboring patches with the local reconstruction weight of 0.975) by Sc, LSc, and LSPSc.**

<https://doi.org/10.1371/journal.pone.0173613.g005>

samples. LSc ensures that sparse representations of global similar samples are consistent, but the local different information among patches is neglected. So it is sensitive to noise and change of background. In contrast, the main contribution of our formulation is the locality preservation, including both the similarity and non-similarity among spatial local patches in images, as illustrated in Figs 5 and 6.

The *Caltech 101* dataset ([http://www.vision.caltech.edu/Image\\_Datasets/Caltech101/](http://www.vision.caltech.edu/Image_Datasets/Caltech101/)), containing 101 categories, is used to demonstrate the principle of our formulation, which is compared with the classical sparse coding Sc (by OMP) and group sparse coding algorithms (LSc). We fix the dictionary size at 1,024;  $\lambda$  0.3, and  $\beta$  0.2. The detailed analysis of parameters can be found in the next section. Template patches are dense grid sampled (template patch number is  $1.5 \times 10^4$ ) from 10 randomly selected images in the *Lotus* category, and  $9.0 \times 10^4$  test patches are sampled from the rest 56 images. All the template patches are used to learn the *lotus* dictionaries by Sc, LSc, and LSPSc. Then, a pair of patches with maximum neighborhood weight of 0.975 selected from test patches are sparse represented by Sc, LSc, and LSPSc according to the corresponding dictionary. The sparse representation of the two samples is one with red and one with green. Fig 5(A), 5(B) and 5(C) correspond to different sparse coding methods, in which higher the red and green coincidence degree indicates that the sparse representation consistency of the two is higher. It can be seen that the sparse features of similar patches by Sc vary a lot; they are very similar by LSc; the similarity between sparse features by LSPSc is not as good as that of LSc, but they are roughly consistent.

To further evaluate the non-similarity preserving ability of LSPSc, we dense grid sample template patches (patch number is  $1.74 \times 10^4$ ) from the 30 randomly selected images in the *Leopards* category,  $9.86 \times 10^4$  test patches from the remaining 170 images in this category. To compare three algorithms in Fig 6, we calculate and plot the spatial local reconstruction weights of original test patches as the similarity between samples in the X-axis, and the local



**Fig 6. Analysis of the similarity preserving performance of Sc, LSc, and LSPSc.** The X-axis is the similarity between original samples, evaluated by local reconstruction weights of spatial neighboring patches. The Y-axis is the similarity between sparse codes, evaluated by local reconstruction weights of the sparse codes corresponding to spatial neighboring samples.

<https://doi.org/10.1371/journal.pone.0173613.g006>

reconstruction weights of their sparse codes as the similarity between sparse features by  $Sc/LSc/LSPSc$  in the Y-axis.

The similarity among original patches is totally destroyed by  $Sc$  in (a). Because kNN (k nearest neighbors) is used to compute the similarity weights from global patches in  $LSc$ , the similarity constraint between highly similar patches is strong and the similarity is preserved well in their sparse codes. But the constraint between patches with weak similarity is ignored, which may lead to dissimilar samples having similar sparse codes on the sample manifold. In particular, although it is a global similarity constraint in  $LSc$  but a local similarity used in (b), it does not impair the performance of  $LSc$  due to the equivalence from local similarity to global similarity.  $LSPSc$  constrains the preservation of the local sparse structure by keeping the spatial neighborhood reconstruction weights among samples into their sparse codes. The consistent neighborhood reconstruction weights in input space and sparse code space indicate that when the weight from neighbor to center patch is big (highly similar patches), the corresponding sparse codes are similar; when weight is small (dissimilar patches) the corresponding sparse codes are dissimilar. So the constraint in  $LSPSc$  contains both similarity and dissimilarity. Therefore, (c) shows a more evident linear trend of the similarity between sparse codes against the similarity between patches. It proves that  $LSPSc$  can preserve the local structure of original samples well.

#### Analysis of the performance in infrared target representation by $LSPSc/K-LSPSc$ .

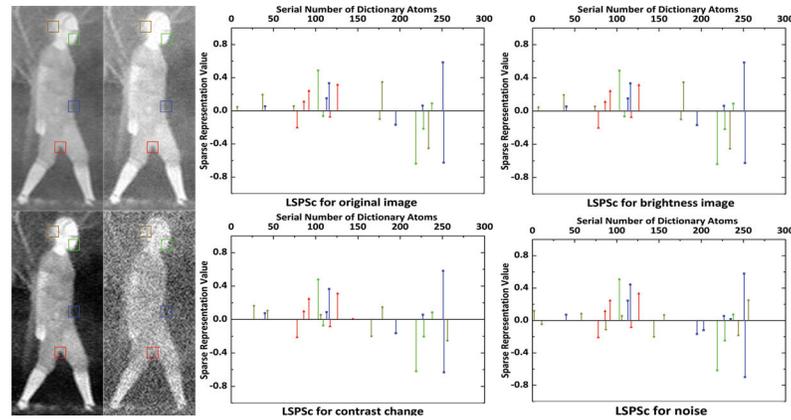
Based on the well property of  $LSPSc$  in visible images, we further use an *Infrared Human* dataset (The *Infrared Human* dataset is captured by FILR Tau 2–640 and Tau 2–324 Uncooled LWIR Thermal Imaging Cameras, including total 224 images with 162 images containing human and 62 images not containing human in different scenes.) to evaluate the performance of  $LSPSc$  and  $K-LSPSc$ . In this dataset, we fix the dictionary size at 256,  $\lambda$  is 0.4, and  $\beta$  is 0.3.

Test the robustness of  $LSPSc$  sparse quantization in infrared image with changing brightness, contrast, and noise. The  $LSPSc$  and  $K-LSPSc$  dictionaries of *infrared human* are learned from the simple template set in Fig 3(A). In Fig 7,  $LSPSc$  sparse representations of brightness changed infrared patches are totally invariant, the contrast changed and noise added infrared patches are slightly influenced. The influence is weak on patches with relatively complex neighborhood structure (green and red), is stronger on patches with simple structure in their neighborhood (blue), and is more prominent on patches with no obvious structure (brown). Because the local sparse structure of target remains unchanged in condition of different environments, more interdependent neighborhood reconstruction relationships indicate a stronger local spatial manifold constrain, which can weaken the interference of noise, imaging blurring and background, and enhance the robustness of sparse quantization in infrared images.

Based on the *infrared human* dictionaries, we measure the local structure preserving by  $LSPSc$  in infrared images. In Fig 8, we present the sparse decoding of the center patch (row 2) and its 8 neighboring patches (row 4) by  $LSPSc$ . When  $\beta = 0$   $LSPSc$  regresses to  $Sc$ . The details in infrared image are fuzzy, stronger constraint of noise and blur invariant structure features can improve the accuracy of sparse decomposition. It is obvious that a bigger  $\beta$  executes a stronger constraint of the spatial local manifold, and the sparse decoding of center patch (row 2) is more similar with the reconstitution from its neighbors (row 3). However, an overlarge  $\lambda$  or  $\beta$  will make it hard for the optimization problem to meet both sparsity and structure preserving simultaneously. This will lead the objective to be nonconvergent or a big convergence value.

By the kernel skill,  $K-LSPSc$  optimizes the linear assumption for actual data in  $LSPSc$ . So under the condition of same parameters,  $K-LSPSc$  achieves more accurate local structure constraints than  $LSPSc$ , and  $K-LSPSc$  has a smaller convergence of average objective value than  $LSPSc$  in Fig 9.

To analyze the identification ability of infrared targets between  $LSPSc$  and  $K-LSPSc$ , we sample subregions from the original image in Fig 10(A), learn the  $LSPSc/K-LSPSc$  matching

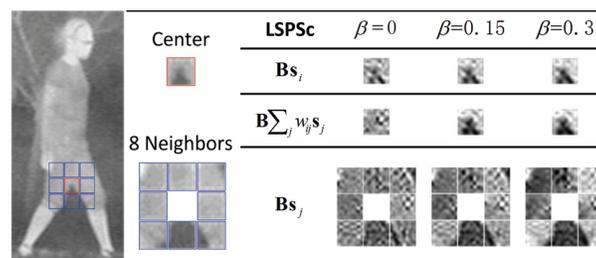


**Fig 7. The invariance and robustness of LSPSc in brightness, contrast, and noise changes.**

<https://doi.org/10.1371/journal.pone.0173613.g007>

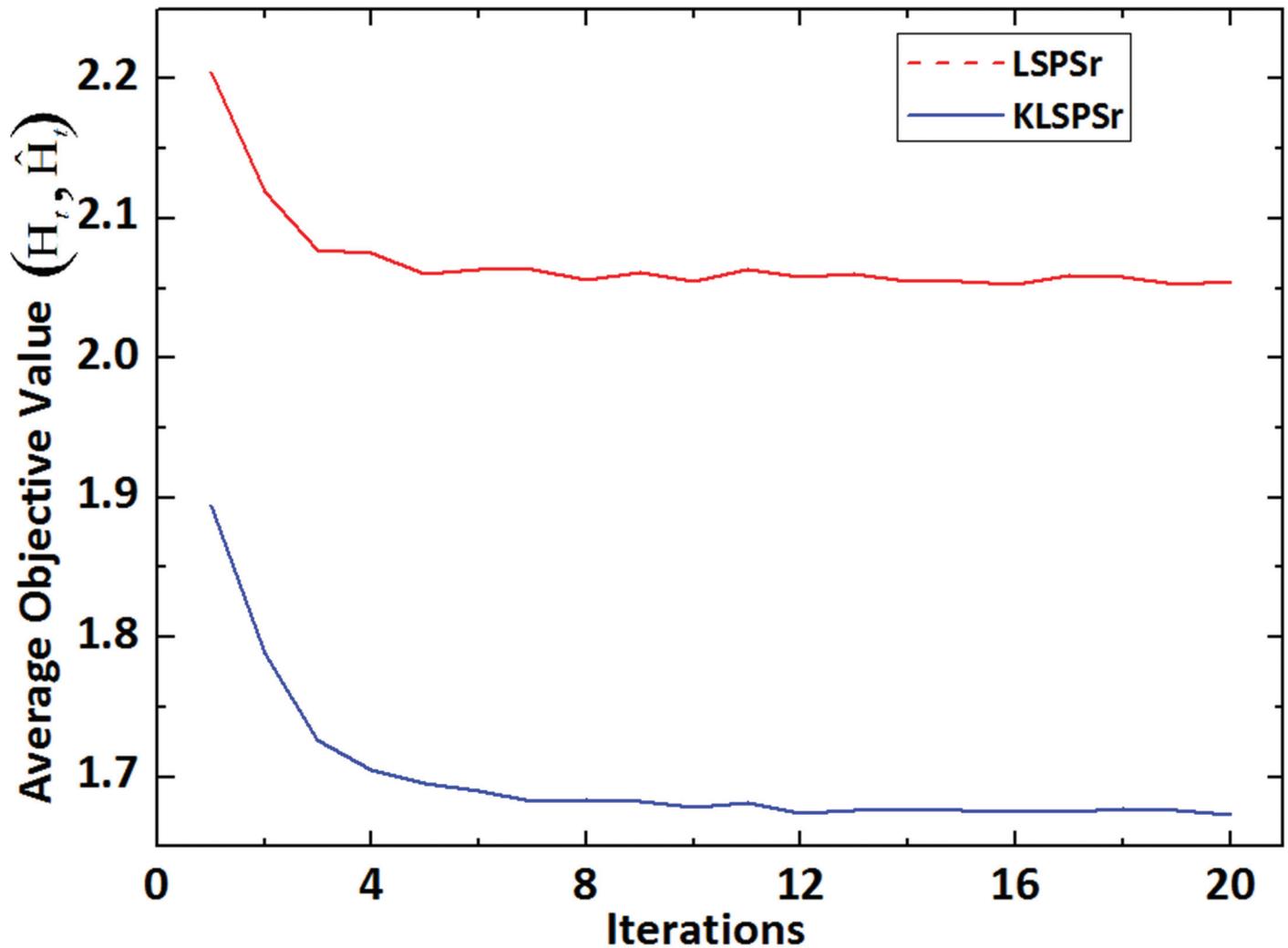
error  $H_t$  by (21) and (22) in each subregion. To facilitate visual comparison, the LSPSc/K-LSPSc matching errors are computed by pixels in the original image. The results are plotted on the right side of Fig 10(A). Based on the human dictionary, the LSPSc/K-LSPSc matching error of pedestrian is small for consistent basic elements and similar structures. Although parts of the basic elements are consistent, the local structures of tree trunk are different from the human body, so the LSPSc/K-LSPSc matching error is large in branch. Therefore, LSPSc/K-LSPSc has the ability to identify infrared target according to the dictionary, and can reduce the influence of noise and fuzzy details in thermal images and natural texture in background. Furthermore, compared with LSPSc, the K-LSPSc matching error is smaller in the areas of human body and homogeneous background, but the relative differences between human and branch is bigger. This is beneficial for K-LSPSc to further reduce the interference of scene and distinguish infrared targets in different categories.

We use the template sets in human, dog, bicycle, and car categories (as instanced in Fig 10(B)) to produce corresponding local sparse structure dictionaries, and count the statistical mean of LSPSc/K-LSPSc matching errors in each category images. The smaller average matching error is, the stronger identification ability of the dictionary to the corresponding object class. First, as a result of the linear method in estimating the spatial local manifold relationship, the target with complex shape (bicycle and human) have a higher average LSPSc matching error, and the target with simple shape (dog and car) have a lower average LSPSc matching error. Second, for targets without subjective postures (bicycle and car), their shape depends only on imaging angles and their K-LSPSc matching error is remarkably lower than that of



**Fig 8. Analysis of spatial local manifold preserving property by LSPSc.** The dictionary size is set at 256,  $\lambda$  is 0.4.

<https://doi.org/10.1371/journal.pone.0173613.g008>



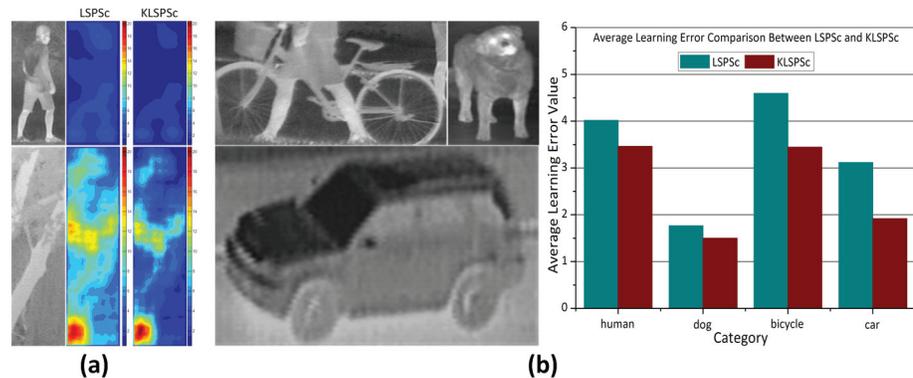
**Fig 9. Comparison of convergence between LSPSc and K-LSPSc.** The average objective value in Y-axis is the mean of LSPSc and K-LSPSc objectives of all patches in Fig 7, whereas the X-axis shows the iterations.

<https://doi.org/10.1371/journal.pone.0173613.g009>

LSPSc. For targets with subjective postures (human and dog), the advantage of K-LSPSc is weak due to the dynamic shapes (changing local structures) of the targets.

For rigid targets, the structural change is mainly caused by the change of imaging angles. The template set itself contains targets in large viewing angle changes, and the same physical structure in small viewing angle changes will produce a certain imaging distortion. LSPSc computation of local nonlinear structure is inaccurate, which makes it difficult to match the different error structures before and after the change of the viewing angle. K-LSPSc is more precise because of the nonlinear calculation of the structure. In a certain range of viewing angles, the patch feature and essential structure in the subregion change little, as shown Fig 11, so the K-LSPSc is more robust to viewing angles. In Table 1, the LSPSc and K-LSPSc objective value  $H_t$  and  $\hat{H}_t$  of red subregions in Fig 11(A), 11(B) and 11(D) are calculated using 11(C) as a template. K-LSPSc results smaller objective values.

To compare the performance in infrared target recognition by KSc, LSc, LSPSc and K-LSPSc-based LSSM, we replace LSPSc/K-LSPSc with the LSc/KSc algorithm, but keep the rest



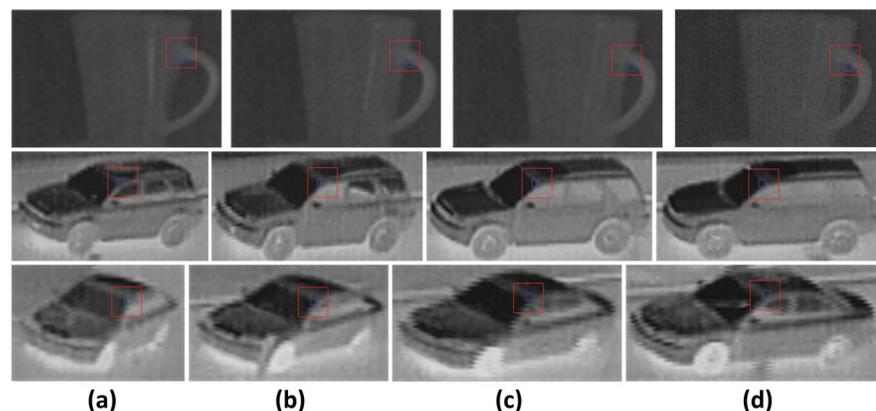
**Fig 10. Comparison of the identification ability to infrared target categories between LSPSc and K-LSPSc.** (a) Original images and their LSPSc/K-LSPSc matching error based on the *infrared human* dictionary; (b) dictionary generation for different categories and comparison of average LSPSc and K-LSPSc matching errors in each category based on the corresponding dictionary.

<https://doi.org/10.1371/journal.pone.0173613.g010>

of the steps in LSSM the same. The extraction of target probability  $\rho_t$  is implemented based on the corresponding *infrared human* dictionaries in the ROIs of original images. ROI can filter out homogeneous background, leaving only target areas consistent with the template. The KSc/LSc/LSPSc/K-LSPSc target probability results are shown in Fig 12B–12E. By requiring only robust sparsity, KSc and LSc are confused by pseudo targets, and the false alarm rates are high (stone, road, and branch in (b) (c)) for some similar local features between pseudo targets and template. There are obviously different neighboring structure relations between pseudo targets and template, which results in a small false alarm rate in LSPSc (d). K-LSPSc has a higher identification ability with more significant peaks of target probability and less background interference than LSPSc (e). Red boxes in (f), same size as the subregion, mark the most potential target areas by the target probability image.

## Results and discussion

To demonstrate the performance of LSSM by using the simple template set, the contrastive analysis of different sparse methods [52] based on a single template and multiple templates is provided. Besides, LSSM is compared with other general target detection method [34,35],



**Fig 11. The differences of patch and subregion feature in changing viewing angles.**

<https://doi.org/10.1371/journal.pone.0173613.g011>

**Table 1. Comparison of LSPSc and K-LSPSc in changing viewing angles.**

Dictionary	Subregions	LSPSc $H_t$	K-LSPSc $\hat{H}_t$
Cup (c)	Cup (a)	0.99	0.87
	Cup (b)	0.95	0.86
	Cup (d)	1.01	0.96
Car1 (c)	Car1 (a)	1.57	1.49
	Car1 (b)	1.49	1.46
	Car1 (d)	1.51	1.47
Car2 (c)	Car2 (a)	1.67	1.58
	Car2 (b)	1.53	1.54
	Car2 (d)	1.62	1.55

<https://doi.org/10.1371/journal.pone.0173613.t001>

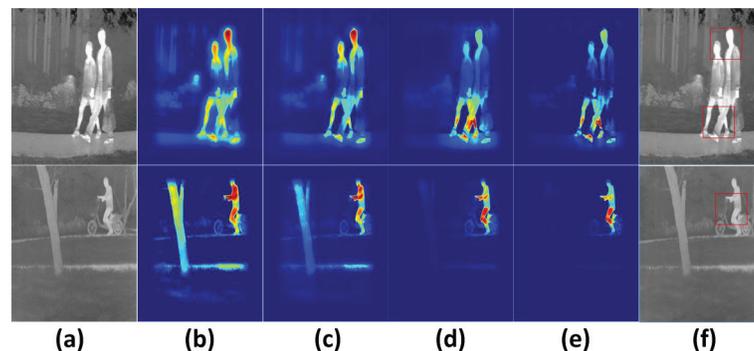
which uses a robust LARK feature to implement entire matching with a single template, and has good capability to detect human faces and targets with simple shape and compact structure.

In this section, experiments are executed on three data sets: *Caltech 256*, *Infrared Car and Bicycle* (captured by FILR Tau 2–640 Uncooled LWIR Thermal Imaging Camera, including total 155 images with 95 images containing bicycle and 60 images not containing bicycle, 87 images containing car and 68 images not containing car.), and *Infrared Human* dataset. For fair comparison, we use the approach in [52] to label bounding boxes around interested targets. If the detected region by our methods lies on the ground truth, we evaluate it as a correct detection and a false positive otherwise. The receiver operating characteristic (ROC) and the Recall versus 1-Precision curves [52] are employed to measure performance.

### Parameter setting

In LSPSc/K-LSPSc formulation there are three free variables: the dictionary size, the weight of sparse term  $\lambda$ , and the weight of local manifold preserving term  $\beta$ . Some works [20,50] have been done to show the relationship from the dictionary size and sparsity to image classification accuracy. We employ the *Caltech 256* and *Infrared* datasets to study the dependence of the target recognition accuracy on three important parameters.

We build two mixed datasets from the *Caltech 256* dataset ([http://www.vision.caltech.edu/Image\\_Datasets/Caltech256/](http://www.vision.caltech.edu/Image_Datasets/Caltech256/)), which contains 256 categories. One is the *Airplanes* dataset,



**Fig 12. Comparison of target probability extraction by KSc, LSc, LSPSc, and K-LSPSc.** (a) the source image; (b–e) the outputs of KSc/LSc/LSPSc/K-LSPSc target probability multiplied with ROIs of original images.

<https://doi.org/10.1371/journal.pone.0173613.g012>

consisting of 750 positive and 750 negative images. The 750 positive images are from the *Airplanes* category (800 images, 50 of them used to learn the *airplanes* dictionary), whereas the 750 negative images are randomly sampled from the rest of the categories in *Caltech 256*. The other dataset is the *Leopards* dataset, consisting of 160 positive and 160 negative images. The 160 positive images are from the *Leopards* category (190 images, 30 of them used to learn the *leopard* dictionary), whereas the 160 negative images are randomly sampled from the rest of the categories in *Caltech 256*. Based on the Sc/LSPSc/K-LSPSc dictionaries of *airplanes*, *leopard*, *infrared car* and *infrared human*, we apply the Sc-, LSPSc- and K-LSPSc-based LSSM on the four datasets, and count the corresponding recognition accuracy, which is defined as  $Acc = (TP + TN)/(nP + nN)$ , where  $nP$  is the total number of positive, and  $nN$  is the total number of negative. Table 2 and Fig 13 show the results.

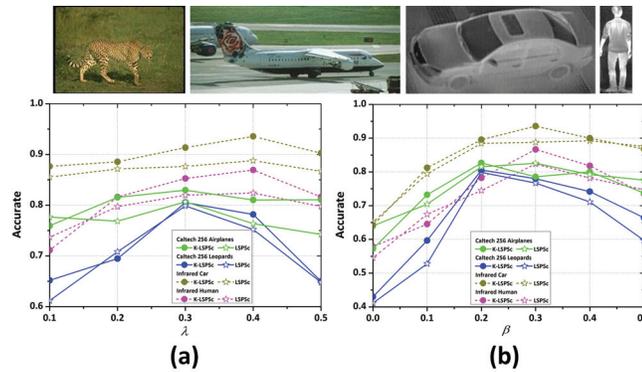
Intuitively, if the dictionary size is too small, LSSM may lose discriminant ability; if the dictionary size is too large, LSSM will be time consuming. As shown in Table 2, for the *Caltech 256 Leopards* dataset, in which the visible targets have rich texture details, the performance of Sc/LSPSc/K-LSPSc increases as the dictionary size goes up to 1024. For the *Infrared Car* and *Infrared Human* datasets, in which the infrared targets have simple structures and blurred textures, the performance of Sc/LSPSc/K-LSPSc peaks when the dictionary size is 256 or 512. In addition, static targets get higher recognition accuracy than dynamic targets (*Airplanes* versus *Leopards*, *Infrared Car* versus *Infrared Human*). Infrared targets achieve higher accuracy than visible objects for the relatively less textures and simple structures (*Infrared Human* versus *Leopards*, *Infrared Car* versus *Airplanes*). It can be seen that for the targets with complex and changing structures, the dictionary size should be large (*Airplanes* and *Leopards*). However, for targets with simple structures, performance has no evident improvement as the dictionary size grows. Therefore, in experiments the dictionary size is set as 1024 for visible dynamic targets and 256 for visible static and infrared targets, to reduce the computed amount.

The weight of sparse term  $\lambda$  enforces the sparsity of the solution. The weight of local structure preserving term  $\beta$  constrains the locality of targets. Fig 13 shows that  $\lambda$  has little effect on recognition accuracy, but the accuracy is impacted greatly by  $\beta$ . Accuracy increases initially and then decreases as  $\beta$  grows further. Overall, LSPSc/K-LSPSc achieve good performance when  $\lambda$  values are 0.3~0.4 and  $\beta$  values are 0.2~0.3. In visible images, excessive structure constrain makes the solution difficult to converge, thus causing the reduction in accuracy. In infrared images, the details are fuzzy and local structures are simple relatively, so appropriate

**Table 2. The effects of dictionary size on Sc, LSPSc, and K-LSPSc on datasets by LSSM.**

Dataset	Dictionary	LSSM methods	Dictionary Size		
			256	512	1024
<i>Caltech 256 Leopards</i>	<i>leopard</i>	Sc	58.48	60.56	<b>61.27</b>
		LSPSc	70.08	76.01	<b>79.84</b>
		K-LSPSc	75.89	77.18	<b>80.52</b>
<i>Caltech 256 Airplanes</i>	<i>airplanes</i>	Sc	62.17	61.41	<b>66.33</b>
		LSPSc	78.16	76.69	<b>80.54</b>
		K-LSPSc	78.82	<b>82.90</b>	81.93
<i>Infrared Car and Bicycle</i>	<i>infrared car</i>	Sc	75.12	<b>76.24</b>	74.57
		LSPSc	<b>88.79</b>	88.54	87.12
		K-LSPSc	<b>93.56</b>	93.47	91.03
<i>Infrared Human</i>	<i>infrared human</i>	Sc	63.16	<b>63.48</b>	62.65
		LSPSc	81.48	<b>82.39</b>	81.91
		K-LSPSc	<b>86.94</b>	85.18	85.27

<https://doi.org/10.1371/journal.pone.0173613.t002>



**Fig 13. The effect of the weight of the sparse term  $\lambda$  and the weight of the structure term  $\beta$  on target recognition accuracy in the *Caltech 256 Leopards, Airplanes, Infrared Car, and Infrared Human* datasets by LSSM.**

<https://doi.org/10.1371/journal.pone.0173613.g013>

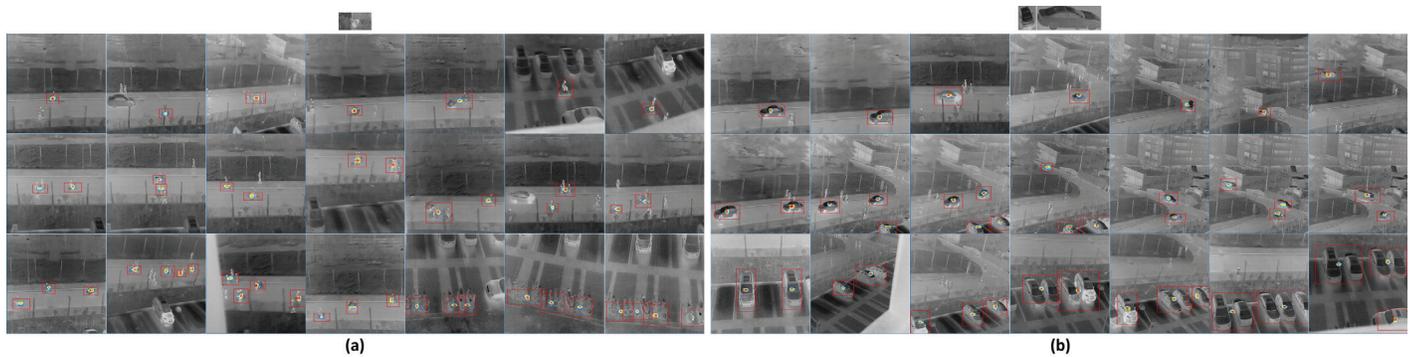
enhancement of the structure constraint can improve accuracy. Thus, for the *Caltech 256* dataset,  $\lambda = 0.3$  and  $\beta = 0.2$ , and for the *Infrared* dataset,  $\lambda = 0.4$  and  $\beta = 0.3$ .

### Infrared vehicle detection

In the *Infrared Car and Bicycle* dataset, the images with size  $640 \times 480$  have little scene changes. Two experiments are conducted with the dataset. One involves 155 images for bicycle and car detection, which contain bicycles and cars from side to front view at various sizes, with a ratio of bicycle size about 2 and a ratio of car size about 3. We use a simple bicycle template set to detect bicycles from 95 positives and 60 negatives, as shown in Fig 14(A). A simple car template set is used to detect cars from 87 positives and 68 negatives, as shown in Fig 14(B). In addition, as explained in Section 3, we construct an expanded template set to deal with the cases of multi-scale, multi-rotation, and different imaging angles of infrared targets. We fix  $\lambda = 0.4$ ,  $\beta = 0.3$ , and the dictionary size at 256. Fig 14 shows the original simple template set of bicycle and car, and outputs of LSPSc-based LSSM for bicycle and car detection. It shows that different vehicle types can be detected for most of the consistent local structures from test targets to the template. Due to the expanded template, targets in different scales, rotations, and imaging angles are recognized accurately.

By computing recall and precision at parameters variations in methods (dictionary size, weights of sparse and local manifold preserving, overall threshold, and confidence level), we conduct two experiments to show the overall performance of LSSM on different templates, and compare it against other methods. Fig 15(A) plots the performance of LSPSc/K-LSPSc-based LSSM in bicycle detection, using the simple bicycle template set in Fig 14(A) and a single template of the side view of electric bicycle in the template set. Here, 1-Precision is indeed 1.0-Precision, a low 1-Precision is desirable since precision should be high, and that the ideal system will have points in the top left of the ROC curve. We can clearly see that the performance is obviously improved by a choice of the template images with different imaging angles, and it is quite consistent with the previous observations that K-LSPSc-based LSSM outperforms LSPSc-based LSSM on both templates, although the advantage is not remarkable by using a single template.

Essentially, the proposed model is a local template matching method based on sparse learning with very few template images. Thus, we not only measure among the LSc/KSc/LSPSc/



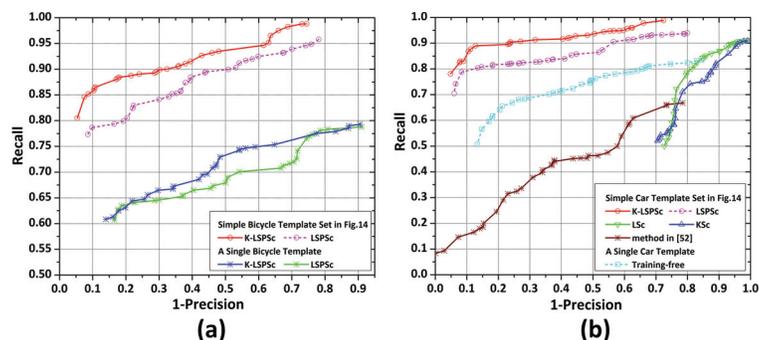
**Fig 14. Recognition results of infrared vehicles (bicycle/car) in different scales, rotations, and imaging angles by LSPSc-based LSSM.** The red boxes are drawn at the target locations.

<https://doi.org/10.1371/journal.pone.0173613.g014>

K-LSPSc-based LSSM models and the part-based sparse representation method [52] by using the simple car template set in Fig 14(B), but also compare LSSM against the training-free method [34,35] by using a single template of the side view of car. Fig 15(B) shows the results of car detection. The method in [52] distinguishes by the similarity between the test samples and “part vocabulary” (constructed by objects of the target class). Because of the imaging characteristics (noise and fuzzy details) in infrared images and use of the simple template set, its similarity discrimination between different car types has low fault-tolerance and adaptability. The training-free method in [34,35] loses the targets in changing imaging angles due to the single template. LSc and KSc mostly pay attention to sparse representation of the test sample, the local features consistent with the template set are detected as a target, so the false positives are high but the false negatives are relatively low. The LSPSc/K-LSPSc-based LSSM keep the consistent local features and sparse structures in neighborhood synchronously, so they have outstanding performance.

### Infrared human recognition

The *Infrared Human* dataset is chosen for human recognition. To be more general to test environments, it is composed of 224 images captured by two thermal cameras in different scenes with resolution ratio of 308\*239, 558\*419 and 617\*506.



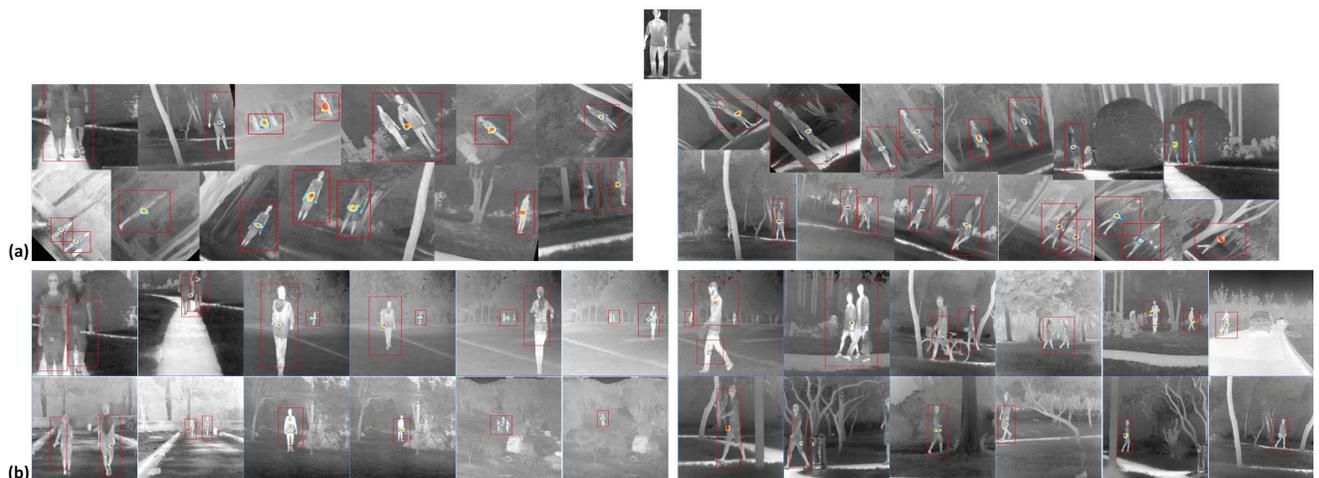
**Fig 15. The performance analysis of LSSM in vehicle detection.** (a) Precision-recall (recall versus 1-precision) curves of bicycle detection by LSPSc/K-LSPSc-based LSSM models using two bicycle templates. (b) Precision-recall curves of car detection by the training-free method using a single car template, method in [52], and LSc/KSc/LSPSc/K-LSPSc-based LSSM methods using a simple car template set.

<https://doi.org/10.1371/journal.pone.0173613.g015>

**Multi-scale and multi-rotation test.** In the vehicle detection we showed the performance of LSSM in the presence of quite different imaging angles but with a moderate scale variation (a size ratio of about 3). In this part, we further evaluate LSSM on a more general scenario for human recognition, where the scale ratio is over 10, various rotations are executed on part images from  $-50^\circ \sim 50^\circ$ , and a large difference of target imaging angles exists. We construct a simple human template set consisting of standing and walking man, to consider both the posture and motion changes of humans. It is expanded in the same manner as the vehicle template set to mainly detect standing and walking humans with slight gesture variances but diverse sizes and rotations. Fig 16 shows the original simple human template set and outputs of LSPSc-based LSSM. Because dictionary learning has programmed the mirroring, scaling and rotation of template images, coupled with a certain ability of anti-interference from background (Fig 10) and fault-tolerant of targets (Fig 12) in LSPSc, LSSM is capable of recognizing and localizing for infrared standing and walking humans based on the few template images as far as possible in Fig 16.

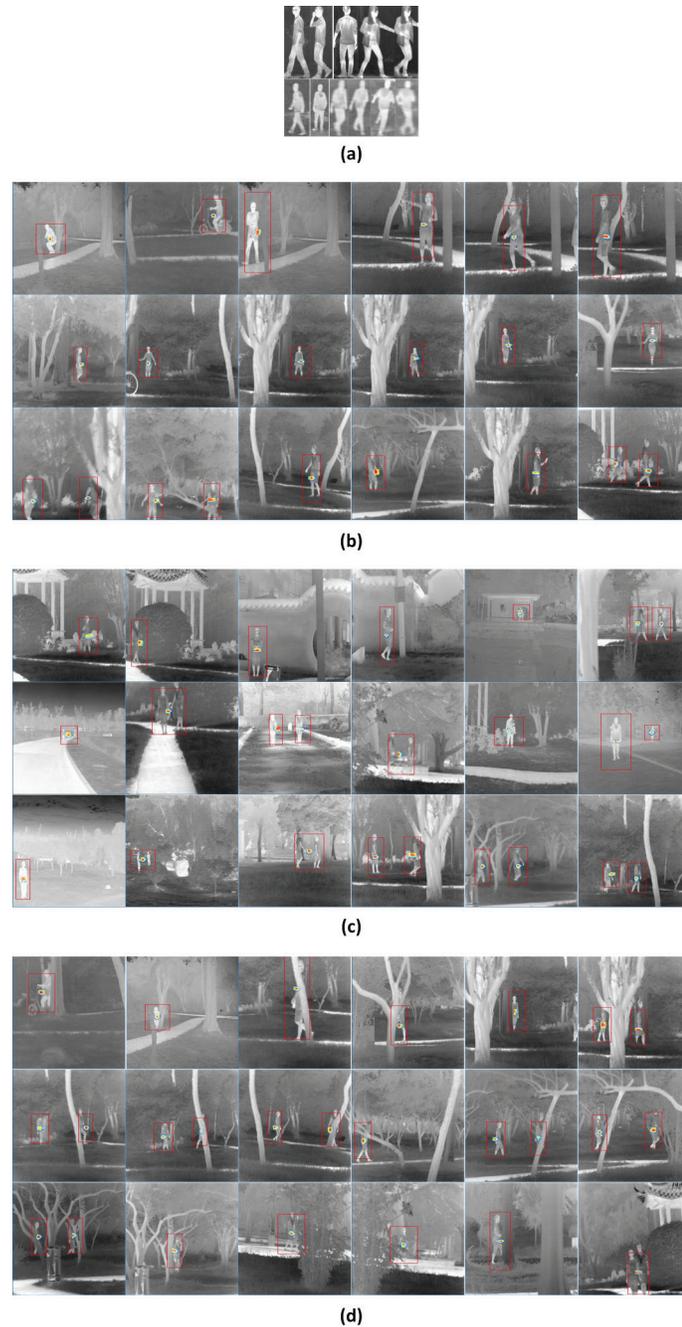
**General human recognition.** Our model is further applied into more complicated scenarios, where we consider robust human recognition in condition of posture variations (Fig 17(B)), scene changes (Fig 17(C)), and occlusion occurrences (Fig 17(D)). To improve the robustness of LSSM, we use the template set in Fig 17(A). It contains more postures and the expansion includes sufficient local features and structures in infrared human body, so the corresponding learned *infrared human* dictionary is relatively robust. Because of the enhanced complication of structures in template and test images, we employ K-LSPSc to further improve the abilities of anti-interference and fault-tolerance in LSSM for identification. Fig 17 shows that the K-LSPSc-based LSSM can recognize and localize the infrared human body reliably in complex environments.

It should be emphasized that, because the *Infrared Human* dataset has targets of standing and walking bodies in different imaging angles and behavior patterns basically, the template set picks several images of standing and walking bodies in representative imaging angles and postures. Moreover, to improve the stability of the dictionary, template images are captured by disparate cameras with different image qualities. If the postures of interested or to be detected humans are abundant, more images of typical motions can be added into the template set further.



**Fig 16. The original simple human template set and results of LSPSc-based LSSM in standing and walking human recognition in large distinctions of scale, rotation, and imaging angle.** (a) multi-rotation; (b) multi-scale. The red boxes are drawn at the target locations.

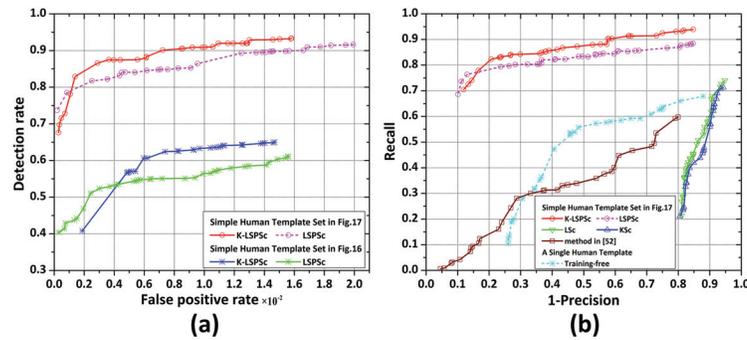
<https://doi.org/10.1371/journal.pone.0173613.g016>



**Fig 17. The robust general infrared human recognition by K-LSPSc based LSSM.** The red boxes are drawn at target locations. (a) Simple template set of human; (b) Robust human recognition with posture variations; (c) Robust human recognition with scene changes; (d) Robust human recognition with occlusions.

<https://doi.org/10.1371/journal.pone.0173613.g017>

Fig 18 shows the ROC curves of LSPSc/K-LSPSc-based LSSM with respect to two different human template sets (template sets in Fig 16 and Fig 17) and recall versus 1-precision curves corresponding to methods in Fig 15(B) (single template is the side view of walking human in Fig 16). It can be seen from Fig 18(A) that the performance of the proposed model is obviously affected by the choice of template sets with different comprehensive degrees of local sparse



**Fig 18. The performance analysis of LSSM in human detection.** (a) ROC curves of LSPSc/K-LSPSc-based LSSM using two human templates; (b) Precision-recall curves of the training-free method using a single human template, method in [52], KSc/LSc/LSPSc/K-LSPSc-based LSSM models using the simple human template set in Fig 17(A). They are recognitions of human body in the *Infrared Human* dataset.

<https://doi.org/10.1371/journal.pone.0173613.g018>

structures. More specifically, in the range of low false positive rates LSPSc provides a higher detection rate, and this phenomenon is stable in the recall versus 1-precision curves in Fig 18(B).

As seen in Fig 18(B), the performance of LSSM in the human dataset is consistent with the results in the vehicle dataset, which detects better than other sparse coding-based methods and the training-free method. As long as the template set contains sufficient local sparse structures in different shapes of the target class, the corresponding LSSM models have stronger robustness for imaging characters, scene variations, shape changes, and partial occlusions of targets. Although the training-free method can recognize human action and the matching process has a certain fault-tolerance, the accuracy using a single template is low in presence of diverse shapes, imaging angles, and occlusions.

### Computation time analysis

Table 3 gives the comparison of computational efficiency on two infrared datasets by different algorithms. It can be seen that LSSM is more time consuming, in particular the KLSPSc-LSSM model. Although the kernel nonlinear processing improves the recognition performance for natural data, it reduces the efficiency of the model. On the other aspect, LSSM training time is obviously higher than training-free method and method in [52]. However, training is usually carried out only once and offline, computational efficiency is more important when testing in real-time implementations, so LSSM has more advantages in the test efficiency and performance.

**Table 3. The comparison of calculational efficiency by different algorithms.**

Dataset	Testing time (s)			
	Training-free method	Method in [52]	LSPSc-LSSM	KLSPSc-LSSM
<i>Infrared Car and Bicycle</i>	9.2411	15.5295	13.8028	30.9442
<i>Infrared Human</i>	13.4669	22.5344	20.0323	44.6790
	Training time (s)			
<i>Infrared Car and Bicycle</i>	0	0.0747	1.5476	2.4638
<i>Infrared Human</i>	0	0.5930	26.9531	49.2764

<https://doi.org/10.1371/journal.pone.0173613.t003>

## Conclusions

By adding a spatial local manifold constraint into the classical sparse coding algorithm, we propose the LSPSc and K-LSPSc formulation in this paper, which can simultaneously preserve both the sparsity of patches and the intrinsic structure of subregions. Moreover, we analyze the relationships between LSPSc/K-LSPSc, Sc, KSc, LSc and structured sparse coding. Experiments show that the local sparse structure quantization by LSPSc/K-LSPSc can alleviate the background interference and improve the stability of sparse representation of infrared targets. We further present an LSSM approach to realize the robust general target detection by using a simple template set. Without plenty of template images, the simple template set of a target class needs only several images containing relatively comprehensive local structures to learn a sufficient sparse structure dictionary.

In the future, our work need to be extended in the following directions: 1) Feature selection: Currently, we use the original gray information in images. How to employ more intrinsic features to improve the discrimination for LSPSc is badly needed. 2) Nonlinear structure constraint: We use a linear manifold to mine the spatial local manifold among samples and embed the local manifold relationships into their sparse quantization. To be appropriate for nonlinear data, LSPSc is extended into kernel space, which entails large computation. We will directly consider a nonlinear manifold method in our future work, such as LTSA, ISOMAP, and so on.

## Supporting information

**S1 File. The source infrared image datasets used in the paper.**  
(ZIP)

## Author Contributions

**Conceptualization:** JH JY.

**Data curation:** JH.

**Formal analysis:** JH JY.

**Funding acquisition:** JH LB.

**Investigation:** JH YZ.

**Methodology:** JH JY.

**Project administration:** JH LB.

**Resources:** JH YZ.

**Software:** JH.

**Supervision:** JH LB YZ.

**Validation:** JH JY.

**Visualization:** JH.

**Writing – original draft:** JH.

**Writing – review & editing:** JH.

## References

1. Chen CP, Li H, Wei YT, Tian X, Tang YY. A Local Contrast Method for Small Infrared Target Detection. *IEEE Transactions on Geoscience and Remote Sensing*. 2014; 52(1): 574–581.

2. Wang Z, Hou QY, Hao L. Improved infrared target tracking algorithm based on mean shift. *Applied Optics*. 2012; 51(21): 5051–5059. <https://doi.org/10.1364/AO.51.005051> PMID: 22858944
3. Gao CQ, Meng DY, Yang Y, Wang YT, Zhou XF. Infrared Patch-Image Model for Small Target Detection in a Single Image. *IEEE Transactions on Image Processing*. 2013; 22(12): 4996–5009. <https://doi.org/10.1109/TIP.2013.2281420> PMID: 24043387
4. Liu HC, Lin ST, Yin HT. Infrared surveillance image super resolution via group sparse representation. *Optics Communications*. 2013; 289: 45–52,.
5. Wright J, Yang AY, Ganesh A, Sastry SS, Ma Y. Robust Face Recognition via Sparse Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2009; 31(2): 210–227. <https://doi.org/10.1109/TPAMI.2008.79> PMID: 19110489
6. Raina R, Battle A, Lee H, Packer B, Ng A. Self-taught learning: Transfer learning from unlabeled data. *Proc. Int'l Conf. Machine Learning*. 2007.
7. Mairal J, Bach F, Ponce J, Sapiro G, Zisserman A. Discriminative Learned Dictionaries for Local Image Analysis. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*. 2008.
8. Zhang Z, Li FZ, S. T, Zhang L, Yan SC. Sparse Codes Auto-Extractor for Classification: A Joint Embedding and Dictionary Learning Framework for Representation. *IEEE Transactions on Signal Processing*. 2016; 64(14): 3790–3805.
9. Li ZC, Liu J, Yang Y, Zhou XF, Lu HQ. Clustering-Guided Sparse Structural Learning for Unsupervised Feature Selection. *IEEE Transactions on Knowledge and Data Engineering*. 2014; 26(9): 2138–2150.
10. Li ZC, Tang JH. Unsupervised Feature Selection via Nonnegative Spectral Analysis and Redundancy Control. *IEEE Transactions on Image Processing*. 2015; 24(12): 5343–5355. <https://doi.org/10.1109/TIP.2015.2479560> PMID: 26394422
11. Li ZC, Liu J, Tang JH, Lu HQ. Robust Structured Subspace Learning for Data Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2015; 37(10): 2085–2096. <https://doi.org/10.1109/TPAMI.2015.2400461> PMID: 26353186
12. Hou CP, Nie FP, Li XL, Yi DY, Wu Y. Joint Embedding Learning and Sparse Regression: A Framework for Unsupervised Feature Selection. *IEEE Transactions on Cybernetics*. 2014; 44(6): 793–804. <https://doi.org/10.1109/TCYB.2013.2272642> PMID: 23893760
13. Zhang Z, Yan SC, Zhao MB. Pairwise Sparsity Preserving Embedding for Unsupervised Subspace Learning and Classification. *IEEE Transactions on Image Processing*. 2013; 22(12): 4640–4651. <https://doi.org/10.1109/TIP.2013.2277780> PMID: 23955747
14. Yan JC, Zhu MY, Liu HX, Liu YC. Face Verification With Local Sparse Representation. *IEEE Signal Processing Letters*. 2013; 20(2): 177–180.
15. Wang JJ, Yang JC, Yu K, Lv FJ, Huang T, Gong YH. Locality-constrained Linear Coding for Image Classification. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*. 2010: 3360–3367.
16. Li XX, Dai DQ, Zhang XF, Ren CX. Structured Sparse Error Coding for Face Recognition With Occlusion. *IEEE Transactions on Image Processing*. 2013; 22(5): 1889–1990. <https://doi.org/10.1109/TIP.2013.2237920> PMID: 23303693
17. Lu ZW, Peng YX. Latent semantic learning with structured sparse representation for human action recognition. *Pattern Recognition*. 2013; 46: 1799–1809.
18. Bai TX, Li YF. Robust visual tracking with structured sparse representation appearance model. *Pattern Recognition*. 2012; 45: 2390–2404.
19. Gao SH, Tsang WI, Ma Y. Learning Category-Specific Dictionary and Shared Dictionary for Fine-Grained Image Categorization. *IEEE Transactions on Image Processing*. 2014; 23(2): 623–634. <https://doi.org/10.1109/TIP.2013.2290593> PMID: 24239999
20. Gao S, Tsang WI, Chia T L. Laplacian Sparse Coding, Hypergraph Laplacian Sparse Coding, and Applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2013; 35(1): 92–104. <https://doi.org/10.1109/TPAMI.2012.63> PMID: 22392702
21. Peng C, Cheng J, Cheng Q. A supervised learning model for high-dimensional and large-scale data. *ACM Transactions on Intelligent Systems and Technology*. 2016; 8(2): 30:1–30:23.
22. Zhao JF, Chen JW, Chen YT, Feng HJ, Xu ZH, Li Q. Sparse-representation-based automatic target detection in infrared imagery. *Infrared Physics & Technology*. 2013; 56: 85–92.
23. Rigamonti R, Brown MA, Lepetit V. Are sparse representations really relevant for image classification. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*. 2011: 1545–1552.
24. Zhang L, Yang M, Feng XC. Sparse representation or collaborative representation: Which helps face recognition. *Proc. IEEE Int'l Conf. Computer Vision*. 2011: 471–478.
25. Gao SH, Tsang WI, Chia LT. Sparse Representation With Kernels. *IEEE Transactions on Image Processing*. 2013; 22(2): 423–434. <https://doi.org/10.1109/TIP.2012.2215620> PMID: 23014744

26. Gao SH, Tsang WI, Chia LT. Kernel Sparse Representation for Image Classification and Face Recognition. *Lecture Notes in Computer Science*. 2010; 63(14): 1–14.
27. Zhang L, Zhou WD, Chang PC, Liu J, Yan Z, Wang T, et al. Kernel Sparse Representation-Based Classifier. *IEEE Transactions on Signal Processing*. 2012; 60(4): 1684–1695.
28. Yin J, Liu ZH, Jin Z, Yang WK. Kernel sparse representation based classification. *Neurocomputing*. 2012; 77: 120–128.
29. Chen Y, Nasrabadi N, Tran T. Hyperspectral Image Classification via Kernel Sparse Representation. *IEEE Transactions on Geoscience and Remote Sensing*. 2013; 51(1): 217–231.
30. Kang CC, Liao SC, Xiang SM, Pan CH. Kernel sparse representation with local patterns for face recognition. *Proc. IEEE Int'l Conf. Image Processing*. 2011: 3009–3012.
31. Yang SY, Wang XX, Yang LX, Han Y, Jiao LC. Semi-supervised action recognition in video via Labeled Kernel Sparse Coding and sparse L1 graph. *Pattern Recognition Letters*. 2012; 33: 1951–1956.
32. Cheng Q, Zhou H, Cheng J, Li H. A minimax framework for classification with applications to images and high dimensional data. *IEEE transactions on pattern analysis and machine intelligence*. 2014; 36(11): 2117–2130. <https://doi.org/10.1109/TPAMI.2014.2327978> PMID: 26353055
33. Seo HJ, Milanfar P. Using local regression kernels for statistical object detection. *Proc. IEEE Int'l Conf. Image Processing*. 2008: 2380–2383.
34. Seo HJ, Milanfar P. Detection of Human Actions from a Single Example. *Proc. IEEE Int'l Conf. Computer Vision*. 2009.
35. Seo HJ, Milanfar P. Training-Free, Generic Object Detection Using Locally Adaptive Regression Kernels. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2010; 32(9): 1688–1704. <https://doi.org/10.1109/TPAMI.2009.153> PMID: 20634561
36. Seo HJ, Milanfar P. Action Recognition from One Example. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2011; 33(5): 867–882. <https://doi.org/10.1109/TPAMI.2010.156> PMID: 20733220
37. Seo HJ, Milanfar P. Static and space-time visual saliency detection by self-resemblance. *Journal of Vision*. 2009; 9(12):1–27.
38. Zoidi O, Tefas A, Pitas I. Visual Object Tracking Based on Local Steering Kernels and Color Histograms. *IEEE Transactions on Circuits and Systems for Video Technology*. 2013; 23(5): 870–882.
39. Lee H, Battle A, Raina R, Ng A. Efficient sparse coding algorithms. *Proc. Conf. Neural Information Processing Systems*. 2007; 19.
40. Donoho D. For Most Large Underdetermined Systems of Linear Equations, the Minimal  $l_1$ -Norm Near-Solution Approximates the Sparsest Near-Solution. Technical Report, Stanford Univ. 2004.
41. Zhang KB, Gao XB, Tao DC, Li XL. Single Image Super-Resolution With Non-Local Means and Steering Kernel Regression. *IEEE Transactions on Image Processing*. 2012; 21(11): 4544–4556. <https://doi.org/10.1109/TIP.2012.2208977> PMID: 22829403
42. Xu ZB, Sun J. Image Inpainting by Patch Propagation Using Patch Sparsity. *IEEE Transactions on Image Processing*. 2010; 19(5): 1153–1165. <https://doi.org/10.1109/TIP.2010.2042098> PMID: 20129864
43. Qiao LS, Chen SC, Tan XY. Sparsity preserving projections with applications to face recognition. *Pattern Recognition*. 2010; 43: 331–341.
44. Yu K, Zhang T, Gong YH. Nonlinear Learning using Local Coordinate Coding. *Proc. Conf. Neural Information Processing Systems*. 2009; 22.
45. Roweis S, Saul L. Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science*. 2000; 290: 2323–2326. <https://doi.org/10.1126/science.290.5500.2323> PMID: 11125150
46. Saul L, Roweis S. Think Globally, Fit Locally: Unsupervised Learning of Low Dimensional Manifolds. *Journal of Machine Learning Research*. 2003; 4: 119–155.
47. Chang H, Yeung DY. Robust locally linear embedding. *Pattern Recognition*. 2006; 39:1053–1065.
48. Pan YZ, Ge SS, Mamun AA. Weighted locally linear embedding for dimension reduction. *Pattern Recognition*. 2009; 42: 798–811.
49. Zhang SQ. Enhanced supervised locally linear embedding. *Pattern Recognition Letters*. 2009; 30: 1208–1218.
50. Yang JC, Yu K, Gong YH, Huang T. Linear Spatial Pyramid Matching Using Sparse Coding for Image Classification. *Proc. IEEE Conf. Computer Vision and Pattern Recognition*. 2009.
51. Devernay F. A Non-Maxima Suppression Method for Edge Detection with Sub-Pixel Accuracy. Technical Report RR-2724. Institut National de Recherche en Informatique et en Automatique. 1995.
52. Agarwal S, Awan A, Roth D. Learning to Detect Objects in Images via a Sparse, Part-Based Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2004; 26(11): 1475–1490. <https://doi.org/10.1109/TPAMI.2004.108> PMID: 15521495