2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

Structured Sparse Coding Method for Infrared Small Target Detection in Video Sequence

Chunwei Yang, Huaping Liu, Shouyi Liao, Shicheng Wang

***Abstract*—In this paper, the infrared small target detection in video sequence is investigated. A collaborative structured sparse coding model which incorporates the $L_{1,2}$ and $L_{2,1}$ regularization terms is proposed to detect the infrared small target in video sequence. Further, online dictionary learning is embedded into the model and temporal information is incorporated to eliminate the clutters and noises. Finally, four simulation datasets are constructed to test the proposed method and the experimental validation shows promising results.**

I. Introduction

INFRARED small target detection is the key technique to Infrared Search and Track System (IRST), in which accuracy and robustness are both needed. Infrared imaging terminal guidance mainly includes small target detection, area target detection and tracking. In the process of terminal guidance, the target are from small to large. If the target is correctly detected and tracked when it is small and faraway, the detection phase can be omitted in the subsequent area target tracking process. Meanwhile, modern warfare requires the weapon system to be able to accurately detect the target under the distance as far as possible in order to get more reaction time for more initiative. As a result, infrared small target detection under complex background is the important precondition for precision guidance. However, the complex battlefield, atmosphere radiation, image system, as well as the characteristics of infrared small target, e.g., no information about shape, size and texture, make the infrared small target detection a troublesome problem.

In general, small target detection methods are divided into two classes: the single frame detection methods and the sequential detection methods. Ref. [1] proposed an infrared patch-image (IPI) model and formulated target detection as an optimization problem of recovering low-rank and sparse matrices, which can be effectively solved using stable principal component analysis. Ref. [2] showed a plausible computational model integrating the robust properties of human visual system using Laplacian scale space theory and optimization method. Ref. [3] developed a kernel-based nonparametric regression method for background prediction and clutter removal. Ref. [4] formulated the infrared small target detection as salient region detection and presents a robust directional saliency-based method. These methods can obtain excellent performance, but due to the low signal-to-clutter ratio (SCR), the performance of detecting small targets in a single image could degrade rapidly. Therefore, sequential detection methods are attracting more and more attention in recent years.

From a view of cognitive science, the image information can be decomposed into two parts [5]: redundancy and saliency. The redundancy denotes the information with high regularities, while the saliency represents the novelty part. In the field of infrared small target detection, as to the whole image, small target can be considered as salient information, which has sparse property. While the background is approximately linearly, which can be treated as redundant information (see Fig. 1 for one example).

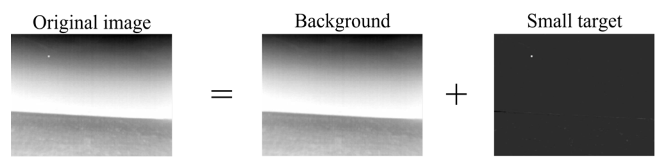

Fig. 1. Infrared image can be decomposed into two parts: background represented as redundancy and small target denoted as saliency.

In this paper, we develop a method called structured sparse coding (SSC) for infrared small target detection in video sequence. The sparse coding methods have been investigated in extensive fields [6][7]. The main contributions of this paper are summarized as follows: (1) A collaborative convex SSC model is proposed to address the infrared small target detection problem. In this model, a $L_{1,2}$ penalty term represents the locations of the infrared small target and $L_{2,1}$ penalty term is used to impose the background. In addition, an iterative optimization method is developed to solve this model. (2) Online dictionary learning and temporal information is incorporated to eliminate the clutters and noises. (3) four simulation datasets are constructed and extensive experimental results are reported to show the role of online dictionary learning and temporal information in infrared small target detection in video sequence.

Notations: Let $\mathbf{M} \in \mathbb{R}^{r \times c}$. We use superscripts for the rows of $\mathbf{M}$, i.e., $\mathbf{M}^{(i)}$ denotes the *i*-th row; and subscripts for the columns of $\mathbf{M}$, i.e., $\mathbf{M}_{(j)}$ denotes the *j*-th column. We will use various matrix norms, here are the notations we use: $\|\mathbf{M}\|_F$ is the Frobenious norm, which is also equal to $\sqrt{Tr(\mathbf{M}^T\mathbf{M})}$; $\|\mathbf{M}\|_{2,1}$ is the sum of the L_2 norm of the rows of

Chunwei Yang is with High-Tech Institute of Xi'an, Xi'an, Shaanxi, 710025, P.R. China, and State Key Laboratory of Intelligent Technology and Systems, TNLIST, Beijing, 100084, P.R.China.

Huaping Liu is with Department of Computer Science and Technology, Tsinghua University, State Key Laboratory of Intelligent Technology and Systems, TNLIST, Beijing, 100084, P.R.China. (corresponding e-mail: hpliu@tsinghua.edu.cn).

Shouyi Liao, Shicheng Wang are with High-Tech Institute of Xi'an, Xi'an, Shaanxi, 710025, P.R.China

This work was supported by the National Key Project for Basic Research of China (2013CB329403).

978-1-4799-1484-5/14/$31.00 ©2014 IEEE

$\mathbf{M}$: $\|\mathbf{M}\|_{2,1} = \sum_{i=1}^{r} \|\mathbf{M}^{(i)}\|_2$; and $\|\mathbf{M}\|_{1,2}$ is the sum of the L_2 norm of the columns of $\mathbf{M}$: $\|\mathbf{M}\|_{1,2} = \sum_{j=1}^{c} \|\mathbf{M}_{(j)}\|_2$.

The remainder of this paper is organized in the following way: In Section II, we overview the SSC method for infrared small target detection in video sequence. In Section III, we present the role of SSC strategy in infrared small target detection. In Section IV, we show the optimization algorithms. In Section V, we give the experimental results and conclusions and perspectives are given in Section VI.

II. Overview of The Proposed Method

Fig. 2 shows the overview of our proposed SSC method for infrared small target detection in video sequence. Firstly, each image $\mathbf{I}$ of the video sequence is partitioned into N blocks of size $a\times b$, producing a set of matrices $D^{(1)}$, $D^{(2)}$, ..., $D^{(N)} \in \mathbb{R}^{p\times N}$, where $p = a\times b$. Secondly, we proceed $\mathbf{D}$=[$D^{(1)}$, $D^{(2)}$, ..., $D^{(N)}$] using SSC strategy, after which we can obtain saliency which denotes the information of the target and noises plus the exemplars which represent the background. Thirdly, we introduce online dictionary learning which means that the main exemplars selected from the exemplars obtained from the t-th frame are embedded into the exemplars of the (t+1)-th frame. By this way, the dictionary is updated online in the whole video sequence. To this end, in order to eliminate noises, we use the temporal information which is inspired by the fact that the target always has a relatively standard trajectory, while the noises are random (illustrated in Fig. 3). Finally, we can obtain the infrared small target from the video sequence.

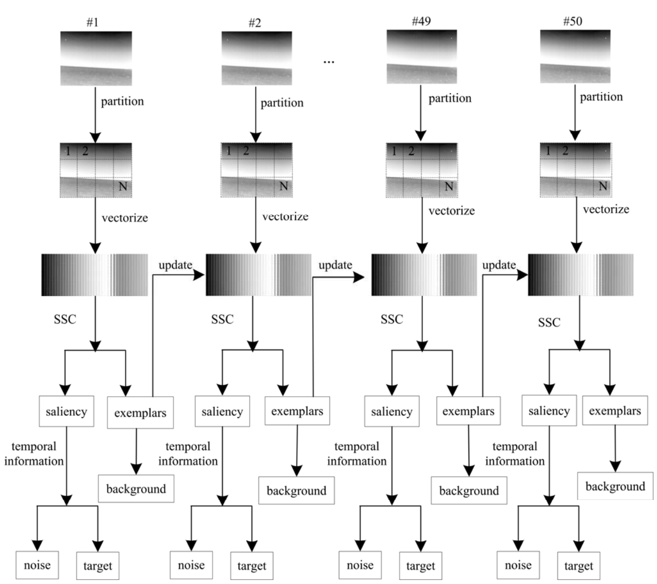

Fig. 2. The overview of the proposed method in this paper. The infrared video sequence is firstly transformed into blocks which are vectorized as a dictionary. Then we obtain exemplars representing background and saliency denoting the target and noises by SSC. After that, selected exemplars are embedded into the next dictionary to achieve the online learning. Meanwhile temporal information is incorporated to eliminate the noise and the location of the real infrared target can be obtained.

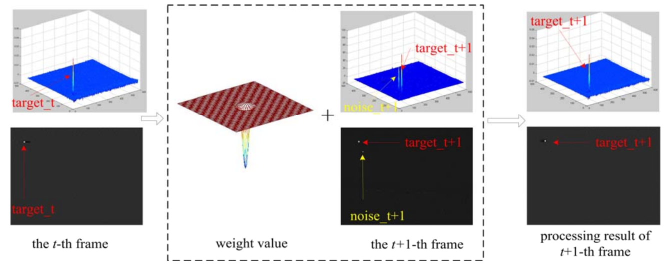

Fig. 3. The illustration of the temporal information embedded in the infrared video sequence. From the t-th frame we can obtain the weight value, which is small in the region of the target and large in other locations. Through the temporal information, the noise can be eliminated and the location of the real target can be obtained.

III. Target Detection Using Sparse Coding

Sparse coding strategy [8]-[11], which is a hotspot in recent years, has been widely used in image denoising, image feature extraction and pattern recognition, etc. As to the infrared small target detection in video sequence, the background can be regarded as a set consists of a collection of representatives extracted from the dictionary, while the target and noises are treated as saliency. Therefore, this problem is equivalent to how to find the saliency denoting the target and select an optimal subset from the dictionary to represent the background under certain constraints.

Consider a matrix $\mathbf{D}_t = \left[D_t^{(1)}, D_t^{(2)}, \cdots, D_t^{(N)}\right] \in \mathbb{R}^{p\times N}$ which is obtained through partitioning the t-th frame of the video sequence, where each column vector denotes a vectorized block. If the infrared small target is in the i-th block, more formally, in column $D_t^{(i)}$, then the corresponding coefficient is sparse. While if the block does not contain the small target, which means it is background, the coefficient will be small and uniform.

Ref. [1] shows that the infrared image contains the small target $\mathbf{T}$, the background $\mathbf{X}$ and the i.i.d random noise $\mathbf{N}$, and noise is thought to be uniform and small. The infrared small target can be obtained by solving the following optimization problem:

$$\min_{\mathbf{X},\mathbf{T}} \|\mathbf{X}\|_* + \lambda\|\mathbf{T}\|_1 + \frac{1}{2\mu}\|\mathbf{D}-\mathbf{X}-\mathbf{T}\|_F^2 \qquad (1)$$

where $\|\bullet\|_*$ is the nuclear norm of a matrix (i.e. the sum of singular values), $\|\bullet\|_1$ is the L_1 norm, λ and μ are weight parameters. However, this model fails when the noise is strong, especially when the noise is similar to the target. What's more, it can be used for single image only.

A. Saliency Detection

In this paper, we consider the noise as sparse and can be treated as saliency. Therefore the task is to find the saliency and then obtain the infrared small target from the saliency. Therefore, a straightforward approach is to minimize the following objective function,

$$\min_{\mathbf{X}_t,\mathbf{E}_t} \|\mathbf{X}_t\|_{2,1} + \lambda\|\mathbf{E}_t\|_{1,2}, \quad s.t. \ \mathbf{D}_t - \mathbf{D}_t\mathbf{X}_t = \mathbf{E}_t \qquad (2)$$

where $\mathbf{X}_t \in \mathbb{R}^{N\times N}$ is the coefficient matrix of the t-th frame and the term $\mathbf{E}_t$ is used to evaluate the saliency of the t-th frame. The parameter λ is used to balance different penalty terms.

The above optimization problem can be efficiently solved [12][13]. After obtaining the value of $\mathbf{E}_t$, the L_2 norm of the j-th column for $\mathbf{E}_t$, which is denoted as $\|\mathbf{E}_{t(j)}\|_2$, is used to evaluate the possibility of the j-th sample to be the saliency of the t-th frame.

In addition, we can also obtain the value of $\mathbf{X}_t$, which is used to evaluate the the possibility of the i-th sample to be the exemplars of the t-th frame.

In such a collaborative structured sparse coding model, as to the t-th frame, the column sparsity is imposed on the matrix $\mathbf{E}_t$ to isolate the saliency and the row sparsity is imposed on $\mathbf{X}_t$ to detect the exemplars. By this way, the saliency is obviously obtained and the extracted exemplars focus on reconstructing the background.

B. Dictionary Update

In order to eliminate the clutters, the dictionary is updated during the detection process. Here, we let $\overline{\mathbf{D}_{t+1}} = [\mathbf{S}_t, \mathbf{D}_{t+1}]$, where $\mathbf{S}_t$ is k representative exemplars selected from the exemplars of the t-th frame in video sequence. Now, we use $\overline{\mathbf{D}_{t+1}}$ instead of $\mathbf{D}_{t+1}$. In this way, the dictionary can be updated online in the whole process.

Therefore, model (2) can be modified as:

$$\min_{\mathbf{X}_{t+1},\mathbf{E}_{t+1}} \|\mathbf{X}_{t+1}\|_{2,1} + \lambda\|\mathbf{E}_{t+1}\|_{1,2}, \quad s.t.\ \mathbf{D}_{t+1} - \overline{\mathbf{D}_{t+1}}\mathbf{X}_{t+1} = \mathbf{E}_{t+1} \tag{3}$$

C. Temporal Weighing

Through processing a number of frames, we can obtain the general information of the trajectory, according to which we design weight value matrices whose entries $\omega_{i,j} = exp(-\beta|e_{i,j}|)$, where $e_{i,j}$ is the reciprocal of the distance to the target and β controls the decaying speed. In this paper, the value of β is empirically set as 0.5.

To embed the temporal information into the model, we extend Eq. (3) as the following:

$$\min_{\mathbf{X}_{t+1},\mathbf{E}_{t+1}} \|\mathbf{X}_{t+1}\|_{2,1} + \lambda\|\mathbf{W}_{t+1}\mathbf{E}_{t+1}\|_{1,2}, \ s.t.\ \mathbf{D}_{t+1} - \overline{\mathbf{D}_{t+1}}\mathbf{X}_{t+1} = \mathbf{E}_{t+1} \tag{4}$$

where $\mathbf{W}_{t+1}$ is a diagonal matrix of the (t+1)-frame, whose diagonal element is $\omega_{i,j}$. The weight $\omega_{i,j}$ reflects the temporal characteristic of the video sequence. In this work we use this strategy to effectively eliminate the strong clutters and noises. Then the real small target can be obtained.

Remark: In general, the main difference between IPI model and our proposed SSC model is the definition of the noise. As to IPI model, the noise is thought to be small and uniform, and is treated as inlier. However, in SSC model, the noise is considered as saliency due to the strong disturbance in the real world. In addition, online dictionary learning and temporal information is incorporated to our SSC model, making it robust to large noises, especially when the noise is similar to the infrared small target.

IV. Optimization Algorithm

The optimization problem in (4) is intrinsically convex and therefore we adopt the Alternating Directional Method of Multiplier (ADMM) [12] to solve Eq. (4). To this end, we transform the above optimization problem as

$$\min_{\mathbf{G}_{t+1},\mathbf{X}_{t+1},\mathbf{E}_{t+1}} \|\mathbf{G}_{t+1}\|_{2,1} + \lambda\|\mathbf{W}_{t+1}\mathbf{E}_{t+1}\|_{1,2}, \ s.t.\ \mathbf{X}_{t+1} = \mathbf{G}_{t+1}, \quad \mathbf{D}_{t+1} - \overline{\mathbf{D}_{t+1}}\mathbf{X}_{t+1} = \mathbf{E}_{t+1} \tag{5}$$

For convenience, we drop the subscripts, and the above equation can be written as

$$\min_{\mathbf{G},\mathbf{X},\mathbf{E}} \|\mathbf{G}\|_{2,1} + \lambda\|\mathbf{WE}\|_{1,2}, \ s.t.\ \mathbf{X} = \mathbf{G}, \mathbf{D} - \overline{\mathbf{D}}\mathbf{X} = \mathbf{E} \tag{6}$$

The augmented Lagrangian associated with the above optimization problem is given by

$$\begin{aligned} L(\mathbf{G},\mathbf{X},\mathbf{E},\mathbf{Y}_1,\mathbf{Y}_2) &= \|\mathbf{G}\|_{2,1} + \lambda\|\mathbf{WE}\|_{1,2} \\ &+ Tr(\mathbf{Y}_1^T(\mathbf{D}-\mathbf{DX}-\mathbf{E})) + \frac{\mu}{2}\|\mathbf{D}-\mathbf{DX}-\mathbf{E}\|_F^2 \\ &+ Tr(\mathbf{Y}_2^T(\mathbf{X}-\mathbf{G})) + \frac{\mu}{2}\|\mathbf{X}-\mathbf{G}\|_F^2 \end{aligned} \tag{7}$$

where $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are dual variables (i.e., the Lagrangian multipliers), μ is a positive scalar. In order to find a minimizer of the constrained problem (6), the ADMM algorithm uses a sequential of iterations

$$\begin{cases} \mathbf{G}^{(k+1)} = \arg\min L(\mathbf{G},\mathbf{X}^{(k)},\mathbf{E}^{(k)},\mathbf{Y}_1^{(k)},\mathbf{Y}_2^{(k)}) \\ \mathbf{X}^{(k+1)} = \arg\min L(\mathbf{G}^{(k+1)},\mathbf{X},\mathbf{E}^{(k)},\mathbf{Y}_1^{(k)},\mathbf{Y}_2^{(k)}) \\ \mathbf{E}^{(k+1)} = \arg\min L(\mathbf{G}^{(k+1)},\mathbf{X}^{(k+1)},\mathbf{E},\mathbf{Y}_1^{(k)},\mathbf{Y}_2^{(k)}) \\ \mathbf{Y}_1^{(k+1)} = \mathbf{Y}_1^{(k)} + \mu(\mathbf{D}-\mathbf{DX}^{(k+1)}-\mathbf{E}^{(k+1)}) \\ \mathbf{Y}_2^{(k+1)} = \mathbf{Y}_2^{(k)} + \mu(\mathbf{X}^{(k+1)}-\mathbf{G}^{(k+1)}) \end{cases} \tag{8}$$

until $\|\mathbf{D}-\mathbf{DX}^{(k+1)}-\mathbf{E}^{(k+1)}\|_F \le \varepsilon$ and $\|\mathbf{X}^{(k+1)}-\mathbf{G}^{(k+1)}\|_F \le \varepsilon$, where ε is the tolerance error. In the following we explain how to solve the optimization problems in (8).

First, the optimization over $\mathbf{G}$ is equivalent to

$$\min_{\mathbf{G}} L(\mathbf{G}) = \frac{2}{\mu}\|\mathbf{G}\|_{2,1} + \|\mathbf{G}-\mathbf{V}\|_F^2 \tag{9}$$

where $\mathbf{V} = \frac{1}{\mu}\mathbf{Y}_2^{(k)} + \mathbf{X}^{(k)}$. According to [13], the i-th row of the optimal solution $\mathbf{G}$ can be analytically obtained as

$$\mathbf{G}^{(i)} = \begin{cases} (1-\frac{1}{\mu\|\mathbf{V}^{(i)}\|_2})\mathbf{V}^{(i)} & \|\mathbf{V}^{(i)}\|_2 > \frac{1}{\mu} \\ 0 & otherwise \end{cases} \tag{10}$$

Secondly, the optimization over $\mathbf{X}$ is equivalent to minimize

$$\begin{aligned} &Tr(\mathbf{Y}_1^{(k)T}(\mathbf{D}-\mathbf{DX}-\mathbf{E}^{(k)})) + \frac{\mu}{2}\|\mathbf{D}-\mathbf{DX}-\mathbf{E}^{(k)}\|_F^2 \\ &+ Tr(\mathbf{Y}_2^{(k)T}(\mathbf{X}-\mathbf{G}^{(k+1)})) + \frac{\mu}{2}\|\mathbf{X}-\mathbf{G}^{(k+1)}\|_F^2 \end{aligned} \tag{11}$$

The solution of $\mathbf{X}$ can be obtained as

$$\begin{aligned} &(\mathbf{I}+\mathbf{D}^T\mathbf{D})^{-1}(\mathbf{D}^T\mathbf{D}-\mathbf{D}^T\mathbf{E}^{(k)}+\mathbf{G}^{(k+1)} \\ &+\frac{1}{\mu}\mathbf{D}^T\mathbf{Y}_1^{(k)} - \frac{1}{\mu}\mathbf{D}^T\mathbf{Y}_2^{(k)}) \end{aligned} \tag{12}$$

Finally, the optimization over $\mathbf{E}$ is equivalent to

$$\min_{\mathbf{E}} L(\mathbf{E}) = \frac{2\lambda}{\mu}\|\mathbf{WE}\|_{1,2} + \|\mathbf{E}-\mathbf{U}\|_F^2, \tag{13}$$

where $\mathbf{U}=\frac{1}{\mu}\mathbf{Y}_1^{(k)}+\mathbf{D}-\mathbf{D}\mathbf{X}^{(k+1)}$. Similar to the solution of **G**, the j-th column of the optimal solution **E** can be analytically obtained as

$$\mathbf{E}_{(j)}=\begin{cases}(1-\dfrac{1}{\mu\omega_j\left\|\mathbf{U}_{(j)}\right\|_2})\mathbf{U}_{(j)} & \left\|\mathbf{U}_{(j)}\right\|_2>\dfrac{\lambda}{\mu\omega_j}\\ 0 & otherwise\end{cases} \quad (14)$$

V. Experimental Results

To evaluate our proposed method, we use the following four baseline methods for comparison: (I) Tophat filtering method [14]. (II) MaxMedian filtering method [15]. (III) MaxMean filtering method [15]. (IV) IPI model [1]. By such a comparison we can clearly show the superior performance of our proposed SSC.

A. Datasets

In order to verify the proposed approach in this paper, four simulation datasets are constructed by using 50 real infrared background images and 1 target and 1 large noise generated artificially. The background images are chosen from a real infrared video sequence. The video sequence in each dataset includes 50 infrared images. The targets in dataset 1 and dataset 2 are in sky and sea, respectively. However, except for adding large noise generated in random location, dataset 3 and dataset 4 are the same as dataset 1 and 2, respectively.

A synthesized image **I** with target and noise can be achieved by embedding a target image **T** and a large noise image **N** with size of $m\times n$ and $p\times q$ respectively into a background image **B**. While the image with target only is constructed in the same way, only to delete the noise synthesized process. Take the former for example. The detail is as follows:

$$\mathbf{I}(x,y)=\begin{cases}\max(\mathbf{T}(x-x_0,y-y_0),\mathbf{B}(x,y)) & \\ \quad x\in(1+x_0,m+x_0), & \\ \quad y\in(1+y_0,n+y_0) & \\ \max(\mathbf{N}(x-x',y-y'),\mathbf{B}(x,y)) & \\ \quad x\in(1+x',p+x'), & \\ \quad y\in(1+y',q+y') & \\ \mathbf{B}(x,y) & otherwise\end{cases} \quad (15)$$

where (x_0,y_0) and (x',y') is the trajectory of the target and a randomly produced pixel location of the noise, respectively, which the left upper corner of the image **I** corresponds to in the image **B** and **N**. Then we blur the synthesized image using Gaussian filter to make it close to a real one. Finally, we obtain dataset 1, 2 without large noises and dataset 3, 4 with large noises. Fig. 4 shows the examples of the synthesized images.

B. Quantitative Evaluation

SCR can be used to describe the difficulty of infrared small target detection. The SCR is defined as follows [16][17]:

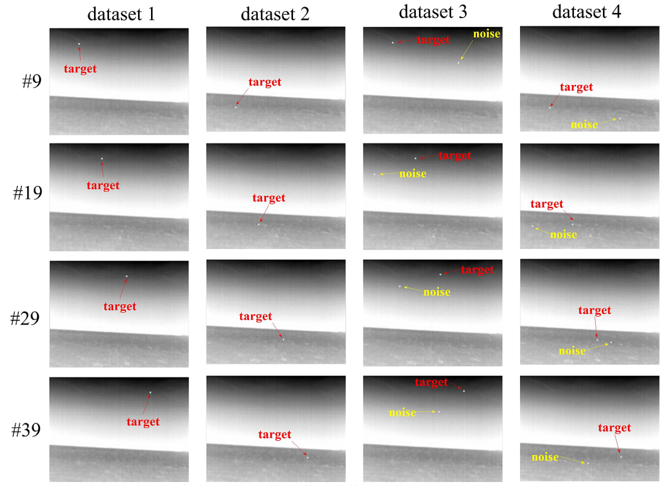

Fig. 4 The selected frames (#9, #19, #29, #39) from dataset 1, dataset 2, dataset 3, dataset 4, respectively. The targets are labeled with red text arrow and the noises are marked with yellow text arrow.

$$SCR=\frac{\mu_t-\mu_b}{\sigma_b} \quad (16)$$

where μ_t is the average value of the target, μ_b and σ_b are the average value and the standard deviation of the pixels of the background, respectively. Base on this, we define the average SCR value of the video sequence as follows:

$$\overline{SCR}=\frac{1}{N}\sum_{t=1}^{N}SCR_t \quad (17)$$

where N is the number of infrared images in video sequence, and SCR_t is the SCR of the t-th infrared image. Also, we define the SCR Gain (SCRG) as:

$$SCRG=\frac{\overline{(SCR)_{out}}}{\overline{(SCR)_{in}}} \quad (18)$$

where $\overline{(SCR)_{in}}$ and $\overline{(SCR)_{out}}$ is the $\overline{SCR}$ of the original and processed infrared images, respectively. In general, the higher the SCRG is, the easier the infrared small target can be detected, which means that the processed method is better.

Table I gives the quantitative evaluation results of the the four baseline methods and our proposed SSC method.

From Table I, we can obviously see that our proposed SSC method is better than the compared ones, and IPI model is better than the other three baseline methods.

C. Qualitative Evaluation

In this section, we list the infrared small target detection results in video sequence of dataset 1 and 2. We show three frames of each processed video sequence in Fig. 5.

From Fig. 5, we can see that the frames processed by IPI model and our SSC method have less clutters compared with others, which means that the infrared small targets of IPI and SSC processed frames are easier to be detected.

D. Robustness to The Large Noise

It is known to us many detection methods in a single image fail when facing noise especially when the noise is large or is similar to the target. In this section, we use dataset 3 and dataset 4 to test the robustness of baseline methods and our proposed SSC. Fig. 6 illustrates the performance of the above

TABLE I
THE PERFORMANCE COMPARISON OF DIFFERENT METHODS OF DATASET 1 AND 2 (THE FILTER SIZE OF TOPHAT, MAXMEDIAN AND MAXMEAN IS 31×31, AND THE PARAMETERS OF IPI MODEL ARE SET AS IN [1])

Methods	*Average_SCR$_{in}$*	*Average_SCR$_{out}$* of dataset 1	*SCRG* dataset 1	*Average_SCR$_{out}$* of dataset 2	*SCRG* dataset 2
Top-hat	0.3625	17.2569	47.6052	10.2849	28.3721
MaxMedian	0.3625	35.7415	98.5972	21.3585	58.9200
MaxMean	0.3625	10.9630	30.2428	4.8470	13.3710
IPI	0.3625	45.3034	127.9749	23.3346	64.3713
SSC	0.3625	46.9213	129.4380	26.6611	73.5479

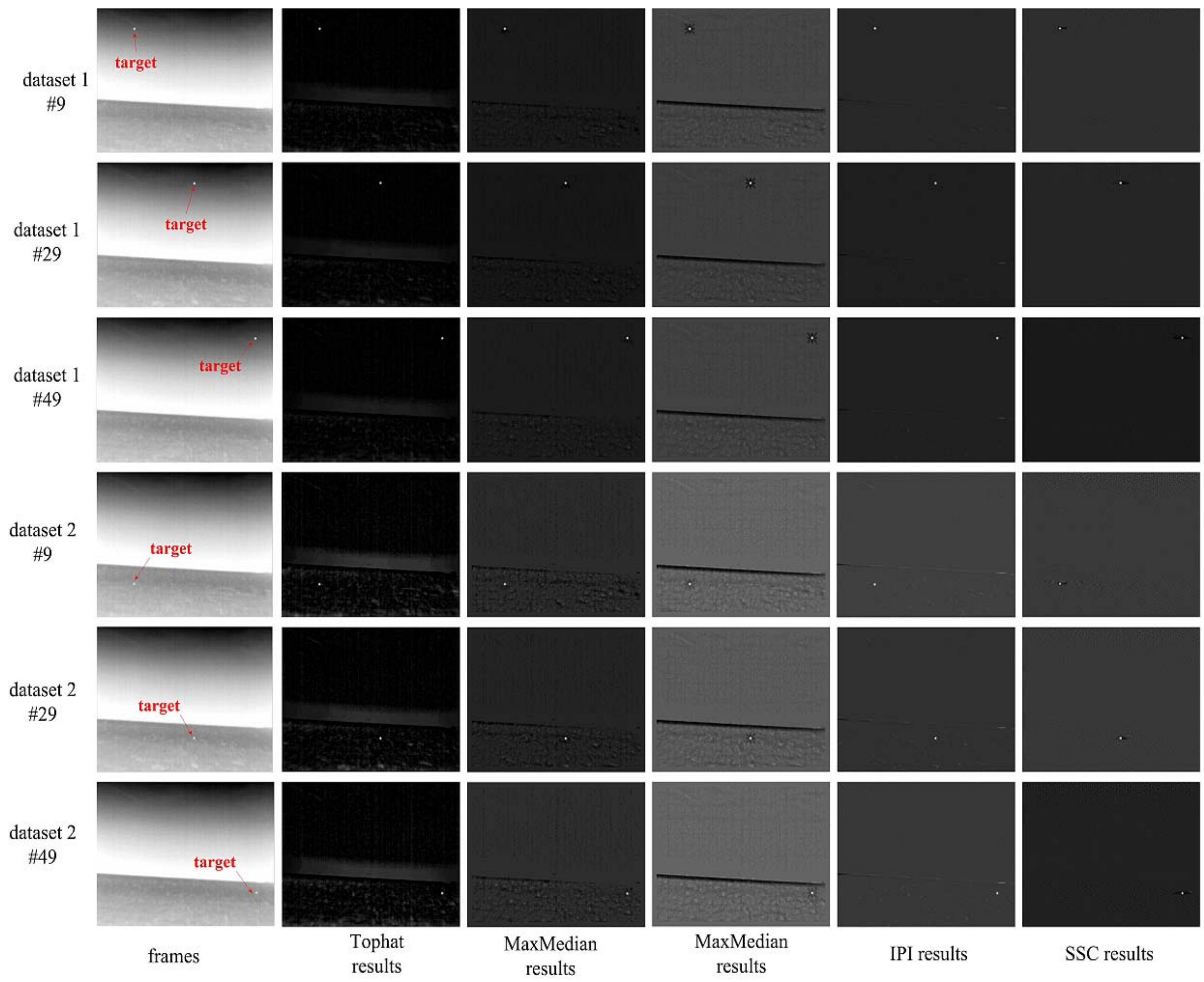

Fig. 5 Frames (#9, #29, #49) of dataset 1 and 2. The first column is frames from datasets and the other five columns are corresponding results proceeded by Tophat, MaxMedian, MaxMean, IPI, and our proposed SSC, respectively.

methods. We can see that the baseline methods is sensitive to the large noise, including the IPI, but our SSC method is very robust to noise.

VI. CONCLUSIONS

In this paper, a collaborative SSC method incorporating the $L_{1,2}$ and $L_{2,1}$ regularization terms is proposed to detect the infrared small target in video sequence. In addition, online dictionary learning and temporal information are embedded into our proposed method to achieve robust capacity. Finally, four datasets are constructed to test the SSC method and the experiments show that the SSC method outperforms those baseline methods, e.g., Tophat, MaxMedian, MaxMean filtering methods and IPI model.

However, there also exists a limitation. As to the proposed SSC model, when the noise near to the target, the robustness

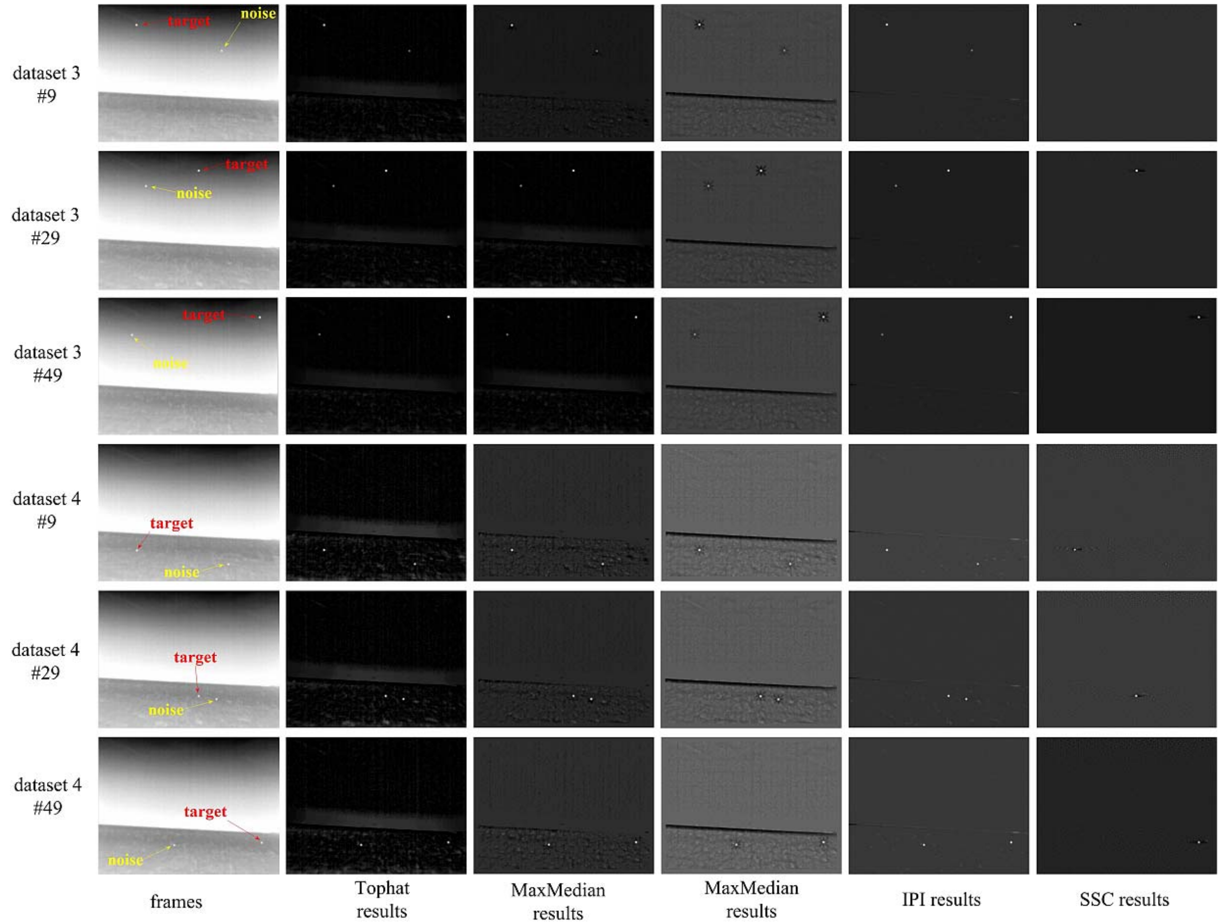

Fig. 6 Frames (#9, #29, #49) of dataset 3 and 4. The first column is frames from datasets from datasets and the left columns are corresponding results proceeded by Tophat, MaxMedian, MaxMean, IPI, and our proposed SSC, respectively.

will get worse. How to solve this problem remains our future work.

REFERENCES

[1] C. Gao, D. Meng, Y. Yang, Y. Wang, X. Zhou, et al, "Infrared patch-image model for small target detection in a single image," *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 4996-5009, Dec. 2013.

[2] S. Kim and J. Lee, "Scale invariant small target detection by optimizing signal-to-clutter ratio in heterogeneous background for infrared search and track," *Pattern Recognition*, vol. 45, issue. 1, pp. 393-406, Jan. 2012.

[3] Y. Gu, C. Wang, B. X. Liu, and Y. Zhang, "A kernel-based nonparametric regression method for clutter removal in infrared small-target detection applications," *IEEE Geoscience and Remote Sensing Letters*, vol. 7, no. 3, pp. 469-473, Jul. 2010.

[4] S. Qi, J. Ma, C. Tao, C. Yang, and J. Tian, "A robust directional saliency-based method for infrared small-target detection under various complex backgrounds," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 3, pp. 495-499, May. 2013.

[5] J. Yan, M. Zhu, H. Liu, and Y. Liu, "Visual saliency detection via sparsity pursuit," *IEEE Signal Processing Letters*, vol. 17, no. 18, pp. 739-742, Aug. 2010.

[6] H. Cheng, Z. Liu, L. Hou, and J. Yang, "Sparsity induced similarity and its applications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. PP, no. 99, pp. 1, Oct. 2012.

[7] H. Cheng, Z. Liu, L. Yang, and X. Chen, "Sparse representation and learning in visual recognition: theory and applications," *Signal Processing*, vol. 93, no.6, pp. 1408-1425, Jun. 2013.

[8] E. J. Candès, X. D. Li, Y. Ma, J. Wright, "Robust principal component analysis?" *Journal of the ACM*, vol. 58, no. 3, pp.11:1-11:37, May. 2011.

[9] S. Yang, Y. Han, X. R. Zhang, "A sparse kernel representation method for image classification," in *World Conference on Computational Intelligence (WCCI)*, Brisbane, 2012, pp. 1-7.

[10] Z. Zhao, C.L. Liu, and M. Zhao. "Handwriting representation and recognition through a sparse projection and low-rank recovery framework," in *International Joint Conference on Neural Networks (IJCNN)*, Texas, 2013, pp. 1-8.

[11] W. E. Vinje and J. L. Gallant, "Sparse coding and decorrelation in primary visual cortex during natural vision," *Science*, vol. 287, no. 5456, pp. 1273-1276, Feb. 2000.

[12] Y. Cong, J. Yuan, and J. Luo, "Towards scalable summarization of consumer videos via sparse dictionary selection," *IEEE Trans. on Multimedia*, vol.14, pp. 66-75, 2012.

[13] E. Elhamifar, G. Sapiro, R. Vidal, "See all by looking at a few: Sparse modeling for finding representative objects," *Proc. Computer Vision and Pattern Recognition (CVPR)*, pp. 1600-1607, 2012.

[14] V. T. Tom, T. Peli, M. Leung, J. E. Bondaryk, "Morphology-based algorithm for point target detection in infrared backgrounds," *Proc. SPIE*, vol. 1954, pp. 25-32, Oct. 1993.

[15] S. D. Deshpande, H. E. Meng, R. Venkateswarlu, and P. Chan, "Max-mean and max-median filters for detection of small targets," *Proc. SPIE*, vol. 3809, pp. 74-83, Oct. 1999.

[16] C. Gao, T. Zhang, Q. Li, "Small infrared target detection using sparse ring representation," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 27, no.3, pp. 21-30, Mar. 2012.

[17] C. I. Hilliard, "Selection of a clutter rejection algorithm for real-time target detection from an airborne platform," *Proc. SPIE*, vol. 4048, pp. 74-84, 2000.