Infrared Physics & Technology 56 (2013) 85–92

Contents lists available at SciVerse ScienceDirect

Infrared Physics & Technology

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

Sparse-representation-based automatic target detection in infrared imagery

Jufeng Zhao, Jinwei Chen, Yueting Chen, Huajun Feng *, Zhihai Xu, Qi Li

State Key Lab of Modern Optical Instrumentation, Zhejiang University, Hangzhou 310027, China

HIGHLIGHTS

- ▶ The theory sparse-representation is used to design method for target detection.
- ▶ Our sparse-representation is based on the idea of center-surrounding difference.
- ▶ The structure similarity is introduced for a stable sparse solution.
- ▶ The method can not only give out the position of targets, but also the whole shape.
- ▶ The approach works accurately in real-life application, gaining low false alarm.

ARTICLE INFO

Article history:
Received 29 April 2012
Available online 22 November 2012

Keywords:
Target detection
Infrared image
Sparse-representation
Center-surround difference

ABSTRACT

A sparse-representation-based automatic target detection approach for forward-looking infrared imagery is proposed. The sparsity is calculated by approximately expressing the pixels within a center image block as the sparsest representation of its surroundings in feature domain. This locally computed sparsity reflects center-surround difference, which is used for distinguishing targets and background clutter. The experimental results on image data show that the proposed approach performs well in IR target detection. The method can not only give out the position of targets, but also the whole shape.

1. Introduction

Research work in forward-looking infrared (FLIR) imagery and automatic target recognition (ATR) has been starting for nearly 40 years. The ATR algorithm usually consists of several stages. According to [1], the first stage is target detection on the entire image; removing background clutter is second stage; in the third and last stages, features are computed and classification is done. In this paper, we mainly focus on the first two stages.

As the fundamental step of ATR, automatic target detection (ATD) technique in FLIR systems has been developed to estimate the location of that target in image. For example, conventional methods such as Max-Mean [2], Max-Median [2], Top-Hat [3] and TDLMS [4], which one can classify as target detection algorithm based on image filtering. Those methods firstly estimate the background, and then subtract this background from original image to get the image with target and noise; at last the target position can be found out with the help with thresholding. There are also other approaches for ATD. A Template matching based algorithm is designed for point targets detection [5]. The detection performance is wonderful. However, the method would fail in extracting the shape of targets. A wavelet multiresolution texture-based algorithm tries to uses the probability density functions (PDFs) of the subband of the wavelet decomposition of an image. The moments of these pdfs are used in a clustering algorithm, with which the targets is segmented from background clutter [6]. Using spatial bilateral filter (BF) and temporal cross product (TCP) of temporal pixels, Bae introduce a spatial and temporal target detection method for infrared (IR) image sequences [7].

Sparse representation has been widely used in image processing, such as face recognition [8,9], signal classification [10]. The idea of Patel et al. [1] is based on the theories of compressive sensing (CS) and sparse representation (SR), which is robust dealing with FLIR images under complicated environment. And Yang et al. [11] defined the similarity vector for clutter metric based on SR (SRC metric). The SR problem finally equals to solve an l_1 minimization problem to gain sparse solution. The sparse solution can provide how closely between the test signal and training samples. There also have been some attempts to use this idea in target detection. Zhao and his partners proposed a small target detection algorithm based on the sparse representation technique [12]. Meanwhile, a joint sparsity model for target detection in hyperspectral imagery is also proposed [13,14]. They both built the SR model by producing a dictionary matrix with training target

* Corresponding author. Tel./fax: +86 571 87951182.
E-mail address: fenghj@zju.edu.cn (H. Feng).

1350-4495/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.infrared.2012.11.001

samples or background samples, which one can consider as prior knowledge. That is to say, they firstly have known the target, and then train it to generate a dictionary matrix, finally use this for target detection with SR model. Their algorithm is similar to the SR based target recognition [1]. However, one usually does not know the target in advance in real-life target detection. How to build the SR model in real-life IR target detection is challenging.

For an excellent target detection algorithm, it can detect the correct targets, including the position and the whole shape, but will not introduce false targets. That is to say, the excellent method should both have high probability of detection (PD) and low probability of false alarms (PFAs). It is a challenging work to design such algorithm, especially only using one image without any prior information. Those common algorithms such as Max-Median [2], Top-Hat [3] and TDLMS [4], can find the position of targets, but not the whole shape. And the PFA is high sometimes, which is one of the disadvantages for target detection. Many excellent algorithms need many image frames, or prior information of targets. However, since one usually cannot know the target information in advance, these prior information based approaches cannot be well used in real-life application sometimes. Moreover, one may only has a few image frames even one image in real-life, those numerous image frames based method cannot work well under this condition.

Inspired by the idea of SR ATR [1] and SRC metric [11], and the concept of target detection using center-surround difference (CSD) [15], in this paper we propose a sparse-representation-based automatic target detection (SRATD) algorithm for IR target images. Combining these previous ideas, we try to seek the optimal solution of SR for the center image block using its surrounding patches. The sparser the solution, the more similar between center and surrounding image blocks. The sparsity is calculated by pixels finally to form a CSD image, which is written as CSD_SR. Since the target object is always different from background, according to which one can detect targets with SR method. One can simply get the location of target with proper threshold for CSD_SR.

2. Sparse-representation-based target detection

If the center block contains a small target, actually, it is difficult to find a sparse solution to make the center block be expressed with the linear combination of its surrounding blocks. Under this condition, the optimal solution f is not sparse, which is we expected. If the optimal solution f is not sparse at certain position, $\|f\|_0$ is relatively large, thus the CSD value is relatively large, and there very likely exists a target. In a word, background patch can be sparsely represented by its neighbors while a target one may not.

This is the rationale for the usage of the sparse-representation assumption in infrared imagery. Therefore, we design the algorithm as follows.

Learning from [1,11,16], we can also briefly describe the design of SR and CS for target detection in infrared imagery. Fig. 1 shows the overview of our approach.

2.1. Sparse representation for our application

This sparse inspired approach is defined as the measurement for difference between target and background. The difference is calculated locally. Fig. 2 can help us understand the definition. The observed signal and dictionary matrix for SR model are designed.

For an arbitrary pixel in the test image, corresponding to the center pixel in Fig. 2, we extend a $t \times q$ block g as the center block which is viewed as an $M \times 1$ column vector, where $M = tq$. As shown in Fig. 2, in order to construct the dictionary matrix, we further extend the $t \times q$ area to $T \times Q$ area, and the surrounding spatial or spatio-temporal image blocks h_i ($i = 1,2,\ldots,N$) is selected with the same size of g. We simply explain how to choose h_i. We define this $T \times Q$ block as B_n, n is the frame number which is corresponding to the 'time' coordinate in Fig. 2. These h_i are allowed overlapping, and cover the entire $T \times Q$ area. Denoting 'row' and 'col' as the coordinates of row and column, $\forall B_n$(row: row + t, col: col + q) $\subset B_n$, $h_i = B_n$(row: row + t, col: col + q), and $h_i \neq g$. For single image, n is fixed, but for image sequences, several adjacent frames of nth frame can be all selected for designing spatio–temporal surrounding blocks h_i.

As $g \in \mathbb{R}^{M\times 1}$, all its surrounding blocks h_i are also viewed as $M \times 1$ column vector, forming an $M \times N$ Dictionary matrix H, $H = [h_1, h_2, \ldots, h_N]$.

We intend to find an optimal $N \times 1$ vector f to make the center signal g be approximately expressed using all the surrounding blocks H, with h_i as its columns:

$$g \simeq \sum_{i=1}^{N} f[i]h_i = Hf \tag{1}$$

The surrounding blocks is allowed overlapping in Fig. 2, resulting in $M \ll N$. It is why we can expect a SR of g. $H \in \mathbb{R}^{M\times N}$ is like a dictionary in [1], which is consist of N surrounding blocks, and the center one g can be approximately represented as a sparse linear combination of those blocks. That implies that the center g can be expressed $g \simeq Hf$, where the vector f has very few nonzero weights. f is a sparse vector.

We try to obtain the sparest representation of g, $g \in \mathbb{R}^N$, one have to solve the Eq. (1). That means one need to solve the following equation as the l_0 norm optimization for f:

$$f_0 = \arg\min_f \|f\|_0 \quad \text{subject to} \quad \|g - Hf\|_2^2 \leqslant \varepsilon \tag{2}$$

where $\|\cdot\|_0$ denotes the l_0 -norm which means the number of nonzero in this vector. And $\|\cdot\|_2$ denotes l_2 -norm, ε is the error tolerance. The more sparse f is, the more similar between center image and surrounding blocks will be. As $M \ll N$, the problem becomes ill-conditioned. The optimization of Eq. (2) is an NP-hard problem, which is computationally difficult to solve. One has to do some approximation. According to [1] [17], as long as f is sparse enough when $M \ll N$ is satisfied, this problem can be solved by minimizing the l_1 -norm:

$$f_0 = \arg\min_f \|f\|_1 \quad \text{subject to} \quad \|g - Hf\|_2^2 \leqslant \varepsilon \tag{3}$$

where $\|f\|_1 = \sum_i |f[i]|$.

2.2. Improved model with constraint

The following problem is the famous least absolute shrinkage and selection operator (LASSO) [18], which is similar to Eq. (3):

$$f_0 = \arg\min_f \|Hf - g\|_2^2 \quad \text{subject to} \quad \|f\|_1 \leqslant \alpha \tag{4}$$

where $\alpha > 0$. By analyzing the optimization of Eqs. (3) and (4), those two constrained optimization can approximate to the unconstrained optimization problem [1]:

$$f_0 = \arg\min_f \left(\frac{1}{2}\|g - Hf\|_2^2 + \lambda\|f\|_1\right) \tag{5}$$

where λ is a regularized factor to balance between the sparsity and distortion, and $\lambda > 0$.

Let $J(f) = \frac{1}{2}\|g - Hf\|_2^2 + \lambda\|f\|_1$, take the derivative of $J(f)$, f_0 becomes the solution of the following equation:

$$\left(\frac{\partial J(f)}{\partial f}\right)\Big|_{f=f_0} = 0 \tag{6}$$