Infrared Physics & Technology 73 (2015) 103–114

Contents lists available at ScienceDirect

Infrared Physics & Technology

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

# Infrared target tracking via weighted correlation filter

Yu-Jie He [a,*], Min Li [a], JinLi Zhang [a,b], Jun-Ping Yao [a]

[a] Xian Research Inst. of Hi-Tech, Xian 710025, China
[b] Engineering University of CAPF, Department of Information Engineering, Xian 710086, China

## HIGHLIGHTS

- An infrared target tracking method consisting of detection and filtering is proposed.
- An infrared target sequential detection algorithm is presented in detection stage.
- A multi-feature weighted function is defined in filtering stage.
- The proposed method yields favorably detection and tracking performance.

ARTICLE INFO

*Article history:*
Received 6 February 2015
Available online 21 September 2015

*Keywords:*
Infrared target detection
Infrared target tracking
Low-rank representation
Weighted correlation filter
Multi-feature weighted function

ABSTRACT

Design of an effective target tracker is an important and challenging task for many applications due to multiple factors which can cause disturbance in infrared video sequences. In this paper, an infrared target tracking method under tracking by detection framework based on a weighted correlation filter is presented. This method consists of two parts: detection and filtering. For the detection stage, we propose a sequential detection method for the infrared target based on low-rank representation. For the filtering stage, a new multi-feature weighted function which fuses different target features is proposed, which takes the importance of the different regions into consideration. The weighted function is then incorporated into a correlation filter to compute a confidence map more accurately, in order to indicate the best target location based on the detection results obtained from the first stage. Extensive experimental results on different video sequences demonstrate that the proposed method performs favorably for detection and tracking compared with baseline methods in terms of efficiency and accuracy.



## 1. Introduction

Infrared target tracking is extremely important, due to its wide range of military applications, such as video surveillance, IR imaging precise guidance and visual supervision. The purpose of infrared target tracking is the same as visual target tracking, that is, to establish the location of the target and track it in subsequent frames. However, in contrast with visual target tracking, infrared target tracking has many more difficult characteristics such as low SNR, non-repeatability of the target signature, lack of shape and textural information, high ego-motion of the sensors and a complicated and chaotic target background. In addition, due to transmission and atmospheric scattering over long observation distances, the targets in sensed videos are usually very dim and small in real environments. All these factors make tracking of infrared targets a more challenging and difficult problem [1–4].

A wide variety of trackers for visual images have previously been proposed, such as L1 [5], CN [6], STC [7], AOG [8], Struck [9], CSK [10] and FCT [11]. However, due to the special IR sequence properties described in the previous paragraph and the differences between visible and infrared target features, only a very limited number of work are suitable for tracking infrared targets. Although many difficulties have been faced, some researchers have made efforts to improve the performance of infrared target tracking, and numerous methods have been proposed in this field. Among these methods, a tracking approach based on detection (tracking-by-detection) has become particularly popular recently [9]. This approach treats the tracking problem as a detection task applied over time. Many detection approaches which have progressed recently can be directly transferred to tracking, and the development of these approaches provides an adaptive tracking mechanism for online classifier training.

* Corresponding author. Tel.: +86 13572962905.
*E-mail address:* ksy5201314@163.com (Y.-J. He).

http://dx.doi.org/10.1016/j.infrared.2015.09.010
1350-4495/

Liu et al. [12] proposed a tracking framework based on template matching combined with Kalman prediction. Such framework used the projection coefficients of PCA as templates and measured the matching degree by using nonlinear correlation coefficients, which constitute the detection phase and provide more precise measurement for Kalman prediction in next tracking step. This method could achieve better performance in infrared targets tracking and was easy to implement, however, it only focused the point and dim targets which can be described by the Gaussian Intensity Function. Dong et al. [13] combined the three mechanisms of Human Visual System (HVS), namely contrast mechanism, visual attention and eye movement, in the proposed detection and tracking approach. Although such method worked well but it was sensitive to the state of the camera, i.e. whether the sensor ego-motion occurs or not. A strategy which could improve the robustness of infrared targets tracking algorithms based on template matching has been presented in [14]. The activation of the template matching, which has a strong impact on tracking performance, is controlled by using a motion prediction metric with reducing the false alarm rate, and it was suitable for the high speed target tracking applications.

Recently, based on the original data is drawn from several low-rank subspaces, Low Rank Representation(LRR) [15] has been proposed for subspace segmentation or recovery, which decomposes the data matrix into a clean matrix and a sparse noise matrix. The clean matrix can be described by self-expressive dictionary with low-rank coefficients. Considering the underlying structure revealing and background modeling ability of LRR with large errors or outliers, the infrared target detection method based on LRR has been proposed [16], however it cannot be used to track infrared targets directly due to the location ambiguity problem. Meanwhile, Average of Synthetic Exact Filters (ASEF) [17] and Minimum Output Sum of Squared Error (MOSSE) [18], et al. have also been proposed for visual tracking, which are both representative algorithms of correlation filter and achieve better tracking performance. In particular, the ASEF and MOSSE have been tremendously successful for eye tracking and obtained the lowest computation complexity by adopting the Fast Fourier Transform. Motivated by these technologies, we present a novel tracking-by-detection method based the weighted correlation filter which is consisting of two parts. For the detection phase, a sequential detection method based on LRR is proposed which improves the detection performance and robustness of LRR. For the filtering phase, we add a multi-feature weighted function to consider the region importance around the target, which combines different features of the detection result to correlation filter named as weighted correlation filter(WCF). The whole tracking process is to firstly obtain the detection results which are used by the WCF to calculate the confidence map and get the most probable target location. Thus, the proposed method has less drift and fuzzy problem as demonstrated by the experiment results.

The rest of the paper is organized as follows. In Section 2, we give a detailed description of the proposed infrared targets detection and tracking method based on the LRR and the correlation filter. The experimental results and the comparisons between the proposed method and the previous classical algorithms are depicted in Section 3. Section 4 provides the conclusions and future works.

## 2. The proposed method

Adaptive tracking-by-detection method, which usually transforms the tracking problem into a classification task which uses learning techniques to update the object model, is a fairly common approach for tracking arbitrary objects. These methods obtain the target locations by searching for the maximum classification score in a local region. However, the objective of tracking is not strictly the same as the classifier objective which is label prediction, and the hypothesis that the highest classifier confidence level corresponds to the best estimate of the object location may cause the tracking results to drift [9]. Therefore, in this paper we propose a different tracking-by-detection approach containing two main conceptual phases, namely target detection phase and filtering phase. The first phase is performed by a proposed sequential detection method based on LRR, and the second phase uses the weighted correlation filter, which calculates the weights using a multi-feature function, to locate the target center. The method process is illustrated in Fig. 1, and the main steps of our approach are as follows.

- Step 1: Patches are extracted using detection windows from $n$ consecutive frames and the detection window group matrix is obtained by adopting the proposed strategy shown as Fig. 2.
- Step 2: The detection window group matrix is decomposed using LRR to obtain the target images and background images. The target image is then post-processed to obtain the final detection result.
- Step 3: The multi-feature weighted function is added into the correlation filter framework. The weighted correlation filter is used to determine the current target location, with the detection result as input.
- Step 4: If it is the last frame, the tracking process stops. Otherwise, it goes back to step 1.

The following subsections provide further details of each part of the proposed tracking-by-detection method.

### 2.1. Target detection based on LRR

Recently, the low-rank recovery problem has become a research hotspot. Considering the given observation matrix was generated from a low-rank matrix $X_0 \in R^{m\times n}$ with a sparse noise $E \in R^{m\times n}$, we could use the following model to recover the original data $X_0$. This model was referred to as robust principle component analysis(RPCA) [19].

$$\min_{Z,E} \operatorname{rank}(X_0) + \lambda\|E\|_l, \; s.t. \; X = X_0 + E \tag{1}$$

where $\lambda$ is a regularization parameter and $\|\cdot\|_l$ indicates a certain regularization strategy. We use the squared Frobenius norm $\|\cdot\|_F$ to model the Gaussian disturbance, $l_0$ norm $\|\cdot\|_0$ to represent the random corruptions, and the $l_{2,0}$ norm $\|\cdot\|_{2,0}$ is used to characterize the sample-specific corruptions (and outliers). However, the RPCA assumes the underlying data is from a single low-rank subspace which dose not fit reality. So the Low Rank Representation (LRR) is formulated as follows [19]. This model considers that data is usually drawn from a union of multiple subspaces in most cases.

$$\min_{Z,E} \|Z\|_* + \lambda\|E\|_l, \; s.t. \; X = AZ + E \tag{2}$$

where $\|\cdot\|_*$ denotes the nuclear norm of a matrix (i.e., the sum of its singular values). $A$ represents a dictionary which spans the union of different subspaces linearly. The optimal solution $Z^*$ indicates the lowest-rank representation of data $X$ according to dictionary $A$. When $A = I$, LRR degenerates to RPCA which can be seen as a special case of LRR model. Usually, we could use observation data $X$ as the dictionary for convenience. So, Eq. (2) becomes:

$$\min_{Z,E} \|Z\|_* + \lambda\|E\|_l, \; s.t. \; X = XZ + E \tag{3}$$