Frontiers of Information Technology & Electronic Engineering
www.zju.edu.cn/jzus; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Detecting slowly moving infrared targets using temporal filtering and association strategy*

Jing-li GAO[1,4], Cheng-lin WEN‡[2,3], Zhe-jing BAO[1], Mei-qin LIU[1]

([1]*College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China*)

([2]*School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China*)

([3]*College of Electrical Engineering, Henan University of Technology, Zhengzhou 450001, China*)

([4]*College of Software Engineering, Pingdingshan University, Pingdingshan 467000, China*)

E-mail: gjl991@163.com; wencl@hdu.edu.cn; zjbao@zju.edu.cn; liumeiqin@zju.edu.cn

Received Apr. 25, 2016; Revision accepted Aug. 9, 2016; Crosschecked Oct. 9, 2016

Abstract: The special characteristics of slowly moving infrared targets, such as containing only a few pixels, shapeless edge, low signal-to-clutter ratio, and low speed, make their detection rather difficult, especially when immersed in complex backgrounds. To cope with this problem, we propose an effective infrared target detection algorithm based on temporal target detection and association strategy. First, a temporal target detection model is developed to segment the interested targets. This model contains mainly three stages, i.e., temporal filtering, temporal target fusion, and cross-product filtering. Then a graph matching model is presented to associate the targets obtained at different times. The association relies on the motion characteristics and appearance of targets, and the association operation is performed many times to form continuous trajectories which can be used to help disambiguate targets from false alarms caused by random noise or clutter. Experimental results show that the proposed method can detect slowly moving infrared targets in complex backgrounds accurately and robustly, and has superior detection performance in comparison with several recent methods.

Key words: Temporal target detection, Slowly moving targets, Graph matching, Target association
http://dx.doi.org/10.1631/FITEE.1601203 **CLC number:** TP391

1 Introduction

Detecting slowly moving targets in infrared (IR) sequences is very useful for many vital tasks, such as low-altitude air defense, air traffic control, and infrared surveillance (Yang *et al.*, 2012; Zhang and Guo, 2012; Wang *et al.*, 2013; Zhang *et al.*, 2013). Since the slowly moving targets acquired from a low-altitude airspace (such as micro-unmanned aerial vehicles (UAVs) and model airplanes) or the targets which are far from the imaging equipment are very dim and small, they do not have available shapes or textures. In addition, the captured targets are usually immersed in heavy noise and clutter, and do not have rich motion features. Although there are some related works in the past decades, the detection of slowly moving targets in heavy noise and clutter is still a challenging problem.

For slowly moving targets in IR sequences, given some prior knowledge of shape and velocity, they can be successfully detected in terms of a target motion model and template matching. However, the effective prior knowledge on targets is hard to obtain in real applications. Furthermore, the sub-pixel velocities computed from several frames are not enough

‡ Corresponding author

* Project supported by the National Natural Science Foundation of China (Nos. 61273170 and 61503206) and the Zhejiang Provincial Natural Science Foundation of China (Nos. LZ16F030002 and LZ15F030001)

ORCID: Jing-li GAO, http://orcid.org/0000-0002-4272-8767

to build precise movement equations. In fact, some targets, such as UAVs and model airplanes, usually fly at a very slow speed. Therefore, detecting slowly moving targets is indeed necessary and is receiving a lot of attention in recent times.

Some conventional methods, such as constant false alarm rate (CFAR) (Zhang *et al.*, 2005), max-mean filter, and max-median filter (Deshpande *et al.*, 1999), pay attention mainly to suppressing the noise and clutter of IR images, and can be used well in the detection of small targets. However, when there exists heavy clutter or evolving clutter in IR images, the detection performance of these methods can severely degrade.

In this study, a detection framework for slowly moving targets is presented. Temporal filtering, temporal target fusion, and cross-product filtering are combined to suppress heavy noise and clutter in IR images first. Then the association strategy is used to associate the obtained detections and suppress false alarms further.

The contributions of this paper are to: (1) propose a temporal target model to suppress heavy noise and clutter, and (2) present a general framework to associate targets and suppress false alarms.

2 Related work

A lot of related methods have been proposed to detect dim targets in the past few decades. These methods can be generally divided into two classes: single-frame based and sequence-based methods. Single-frame based methods usually use the spatial information within a single frame to detect small and dim targets. Gao *et al.* (2013) presented a patch-image model to detect small targets based on low-rank decomposition, which could detect dim targets in fast-moving IR backgrounds. Liu *et al.* (2013) proposed an infrared point target detection method based on template matching and Kalman prediction. The Kalman filter was used to reduce the target searching region, and template matching, based on principal component analysis (PCA), provided more accurate measurements for Kalman prediction. Chen *et al.* (2014) proposed a simple small target detection algorithm based on the contrast mechanism of a human visual system and the derived kernel model. The algorithm uses a local contrast measure to obtain the local contrast map, and then adopts an adaptive threshold to segment the target. Li *et al.* (2014) presented a real-time infrared target detection method under complex conditions. Compressive and sparse features were combined to obtain effective results. Dong *et al.* (2014) used three mechanisms of a human visual system to detect infrared dim targets. A group of difference of Gaussian filters was used to simulate the contrast mechanism and filter the input image, and then a Gaussian window was added at the attention point of the dim target. Finally, the proportional-integral-derivative (PID) algorithm was used to predict the attention point of the next frame. Because of using multiple features, this method can detect dim targets well.

Besides single-frame based methods, many sequence-based methods have also attracted attention. Silverman *et al.* (1996) and Zhang *et al.* (2010) developed a temporal filter, originally suggested by the singular value decomposition (SVD) of consecutive images, to detect slow weak targets in an evolving cloud clutter. Wang *et al.* (2006) fused the gradient features of consecutive frames to suppress residual clutter and reduce false alarms. Kim *et al.* (2014) developed a temporal contrast filter to detect supersonic small infrared targets, and hysteresis threshold-based detection was used to further enhance the accuracy of the target position. Gao *et al.* (2015) enhanced the slow dim target through temporal superposition of consecutive residual images. Liu *et al.* (2015) proposed a moving target detection approach, and employed a nonlinear adaptive filter to remove large fluctuations on temporal profiles. Deng *et al.* (2016) combined a spatial local contrast filter and a temporal local contrast filter to detect the infrared moving point target. Chen *et al.* (2016) proposed a motion saliency detection method, using a temporal Fourier transform to detect the variation in the phase spectrum of consecutive color images. However, these methods could result in false alarms when detecting dim targets with a low signal-to-clutter ratio (SCR).

Furthermore, some association methods based on multiple features of consecutive images were presented. Taj *et al.* (2006) proposed an object detection and tracking algorithm based on color change detection and multiple feature graph matching. Yan *et al.* (2012) presented an ensemble framework for multi-target tracking. This framework computes association scores by discriminative learning and

multi-feature integration, and then uses these scores to associate the tracked targets to the outputs of the trackers and detectors. However, for most infrared small targets, some of the features are unavailable, such as concrete appearance and velocity, and thus these methods are ineffective.

3 Target detection using temporal filtering and association strategy

Our detection approach is illustrated by the framework diagram in Fig. 1, containing two parts: (1) suppression of the complex background based on temporal max filtering and temporal median filtering, enhancement of the targets by the temporal target fusion scheme, and reduction of false alarms based on cross-product filtering; (2) association of the temporal detections, and suppression of the false alarms further based on a graph matching model and a target association strategy.

3.1 Temporal target detection

Let $\boldsymbol{F} \in \mathbb{R}^{M\times N\times L}$ be an IR image sequence, where L is the number of frames and $M \times N$ the size of each frame. A temporal profile of $\boldsymbol{F}$ at pixel position (x, y) is described as $\{\boldsymbol{F}(x, y, k),\ k = 1,\ 2,\ \ldots,\ L\}$. So, there are $M \times N$ temporal profiles in total. Moving a slide window along a temporal profile, the median value (Liu *et al.*, 2015) and the maximum value in the slide window are obtained respectively as follows:

$$\boldsymbol{F}_{\mathrm{med}}(x, y, k) = \mathrm{median}\ \{\boldsymbol{F}(x, y, k + n)\}_{n=-l}^{l}, \quad (1)$$

$$\boldsymbol{F}_{\max}(x, y, k) = \max\ \{\boldsymbol{F}(x, y, k + n)\}_{n=-l}^{l}, \quad (2)$$

where l is an even number. So, the width of the slide window is $2l + 1$.

After obtaining the outputs of the median and maximum filters in terms of Eqs. (1) and (2), we can extract the target pixels as follows:

$$\boldsymbol{F}_{\mathrm{med}}^{\mathrm{t}}(x, y, k) = \boldsymbol{F}(x, y, k) - \boldsymbol{F}_{\mathrm{med}}(x, y, k), \quad (3)$$

$$\boldsymbol{F}_{\max}^{\mathrm{t}}(x, y, k) = \boldsymbol{F}_{\max}(x, y, k) - \boldsymbol{F}(x, y, k). \quad (4)$$

To enhance the targets and further suppress noise, we adopt a fusion scheme to use the temporal information of slowly moving targets. The fusion scheme can be described as (Wang *et al.*, 2006; Gao *et al.*, 2015)

$$\boldsymbol{R}_{\mathrm{med}}^{i}(x, y) = \sum_{k=i-m+1}^{i} \boldsymbol{F}_{\mathrm{med}}^{\mathrm{t}}(x, y, k), \quad (5)$$

$$\boldsymbol{R}_{\max}^{i}(x, y) = \sum_{k=i-m+1}^{i} \boldsymbol{F}_{\max}^{\mathrm{t}}(x, y, k), \quad (6)$$

where i represents the ith image and m is the number of images in which the target information is required to be superimposed.

Calculating the temporal fusion values for all positions (x, y) according to Eqs. (5) and (6), we can obtain the enhanced target images $\boldsymbol{R}_{\mathrm{med}}^{i}$ and $\boldsymbol{R}_{\max}^{i}$.

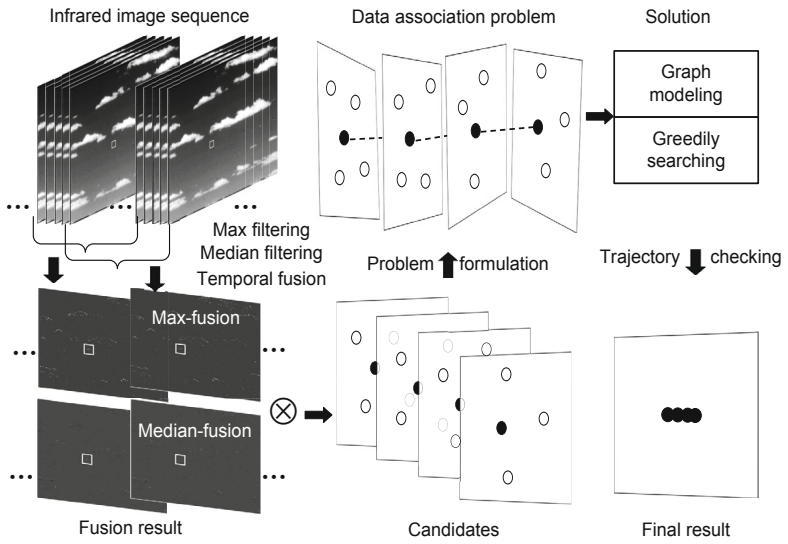

Fig. 1 Framework diagram of the proposed method

Then SVD is used to remove small errors in $\boldsymbol{R}^i_{\text{med}}$ and $\boldsymbol{R}^i_{\max}$. So, the reconstructed images $\widehat{\boldsymbol{R}}^i_{\text{med}}$ and $\widehat{\boldsymbol{R}}^i_{\max}$ can be obtained.

Since the false alarms in $\widehat{\boldsymbol{R}}^i_{\max}$ are chiefly resulted from cloud edges and random noise, and yet those in $\widehat{\boldsymbol{R}}^i_{\text{med}}$ are caused mainly by random noise, a cross-product filter can be defined to suppress false alarms:

$$\boldsymbol{R}^i_{\text{p}}(x,y)=\widehat{\boldsymbol{R}}^i_{\text{med}}(x,y)\times\widehat{\boldsymbol{R}}^i_{\max}(x,y). \tag{7}$$

Repeat the above process, a sequence of target frames $\{\boldsymbol{R}^i_{\text{p}}\}$ can be obtained. Thus, the candidates $\{D^i\}$ can be extracted from $\{\boldsymbol{R}^i_{\text{p}}\}$. The whole process of temporal target detection is described in Algorithm 1.

Algorithm 1 Temporal target detection

Input: $\boldsymbol{F}\in\mathbb{R}^{M\times N\times L}$
Output: $\{D^i\}$
1: **for** each i **do**
2: **for** each (x,y) **do**
3: Calculate $\boldsymbol{R}^i_{\text{p}}(x,y)$ using Eq. (7)
4: **end for**
5: Extract candidates D^i from $\boldsymbol{R}^i_{\text{p}}$
6: **end for**

3.2 Temporal target association

Although the proposed temporal detection method can improve the detection performance, the problem of false detections still exists. Unfortunately, the existence of false alarms could influence subsequent tasks (e.g., decision-making). To solve this problem, we propose a framework for target association and false alarm inhibition based on bipartite graph matching and the temporal detection results.

3.2.1 Graph matching model

Let $D^i=\{\boldsymbol{d}^i_1,\ \boldsymbol{d}^i_2,\ \ldots,\ \boldsymbol{d}^i_r\}$ be a set of r candidates obtained using Algorithm 1. A trajectory $\boldsymbol{T}_j$ is defined as a sequence of well-ordered points $\{\boldsymbol{d}^{i_1}_{c_1},\ \boldsymbol{d}^{i_2}_{c_2},\ \ldots,\ \boldsymbol{d}^{i_n}_{c_n}\}$ such that $l+m\le i_1<i_2<\ldots<i_n\le L-l$ and $1\le c_1<c_2<\ldots<c_n\le r$. Given the detections D^i and the set of trajectories $\mathcal{T}^{i-1}=\{\boldsymbol{T}^{i-1}_1,\ \boldsymbol{T}^{i-1}_2,\ \ldots,\ \boldsymbol{T}^{i-1}_s\}$ obtained at time t_{i-1}, the target association problem is formulated as an assignment problem that associates the detections D^i with the set of trajectories $\mathcal{T}^{i-1}$.

Let $\xi^{i-1}=\{\boldsymbol{e}^{i-1}_1,\ \boldsymbol{e}^{i-1}_2,\ \ldots,\ \boldsymbol{e}^{i-1}_s\}$ be the set of end points of the trajectories in $\mathcal{T}^{i-1}$; these points are selected from detections D^{i-1} or other former detections, and each trajectory is responsible for one target. Denote $G=(V,\ E)$ an edge weighted bipartite graph, where V and E are the sets of vertices and edges of G. Given that $V=D^i\bigcup\xi^{i-1}$ $(D^i\bigcap\xi^{i-1}=\oslash)$, the assignment problem for detections D_i is modeled by the bipartite graph G with an association score as follows:

$$\begin{aligned}&\arg\max_{\{\boldsymbol{d}^i_p\}}\sum_{j=1}^{k}\sum_{w=1}^{4}\alpha_w\cdot\Psi_w(\boldsymbol{e}^{i-1}_j,\ \boldsymbol{d}^i_p)\\&\text{s.t. }\forall(\boldsymbol{e}^{i-1}_a,\ \boldsymbol{d}^i_u)\text{ and }(\boldsymbol{e}^{i-1}_b,\ \boldsymbol{d}^i_v)\ u\ne v\text{ if }a\ne b,\\&\boldsymbol{e}^{i-1}_j,\ \boldsymbol{e}^{i-1}_a,\ \boldsymbol{e}^{i-1}_b\in\xi^{i-1},\ \boldsymbol{d}^i_p,\ \boldsymbol{d}^i_u,\ \boldsymbol{d}^i_v\in D_i,\end{aligned} \tag{8}$$

where $\Psi_w(\boldsymbol{e}^{i-1}_j,\ \boldsymbol{d}^i_p)$ denotes the association gain between the detections centered at point $\boldsymbol{e}^{i-1}_j$ and detection $\boldsymbol{d}^i_p$, α_w is the corresponding weighting parameter, and the sum of all weighting parameters is one. We can compute four types of gains based on the target position, appearance, energy, and size. The formulation in Eq. (8) tries to assign one detection to at most one trajectory such that the matching between detections and trajectories attains the optimum. The association gain for each combination $(\boldsymbol{e}^{i-1}_j,\ \boldsymbol{d}^i_p)$ is described below in detail.

Given position $\boldsymbol{e}^{i-1}_j$ and detection $\boldsymbol{d}^i_p$, the position gain Ψ_1 is computed as

$$\Psi_1(\boldsymbol{e}^{i-1}_j,\ \boldsymbol{d}^i_p)=1-\frac{\text{dist}(\boldsymbol{e}^{i-1}_j,\ \boldsymbol{d}^i_p)}{\rho}, \tag{9}$$

where ρ is a normalizing constant and dist() denotes the Euclidean distance between two points. The appearance gain Ψ_2 is the Bhattacharyya distance between the gray histogram $\hat{p}$ of detection $\boldsymbol{d}^i_p$ and the gray histogram $\hat{q}$ of the detection centered at $\boldsymbol{e}^{i-1}_j$, and can be computed as (Comaniciu *et al.*, 2000)

$$\Psi_2(\boldsymbol{e}^{i-1}_j,\ \boldsymbol{d}^i_p)=\sum_{w=1}^{l_2}\sqrt{\hat{p}_w\hat{q}_w}, \tag{10}$$

where $\hat{p}=\{\hat{p}_w\}_{w=1}^{l_2}$ and $\hat{q}=\{\hat{q}_w\}_{w=1}^{l_2}$. The geometric interpretation of Ψ_2 is the cosine of the angle between histograms $\hat{p}$ and $\hat{q}$. The larger the value of Ψ_2, the larger the similarity between detection $\boldsymbol{d}^i_p$ and the detection centered at $\boldsymbol{e}^{i-1}_j$. Since the energy of infrared targets does not involve a dramatic change in a short span of time and is another efficient

feature, it can be used to calculate the energy gain Ψ_3:

$$\Psi_3(\boldsymbol{e}_j^{i-1},\ \boldsymbol{d}_p^i) = \sum_{(x,y)\in A_e} \widehat{\boldsymbol{R}}_{\max}^{i-1}(x,y) - \sum_{(x,y)\in A_d} \widehat{\boldsymbol{R}}_{\max}^{i}(x,y), \tag{11}$$

where A_e is a region of 3×3 centered at $\boldsymbol{e}_j^{i-1}$, A_d is also a region of 3×3 which has the same center point as $\boldsymbol{d}_p^i$, and x, y denote the coordinates of a point in A_e or A_d. Suppose that the size of the target centered at $\boldsymbol{e}_j^{i-1}$ is $W_e^{i-1}\times H_e^{i-1}$, and that the size of detection $\boldsymbol{d}_p^i$ is $W_d^i\times H_d^i$. So, the size gain Ψ_4 can be defined as (Taj *et al.*, 2006)

$$\Psi_4(\boldsymbol{e}_j^{i-1},\ \boldsymbol{d}_p^i) = 1-\frac{1}{2}\left(\frac{|W_e^{i-1}-W_d^i|}{\max(W_e^{i-1},W_d^i)} + \frac{|H_e^{i-1}-H_d^i|}{\max(H_e^{i-1},H_d^i)}\right). \tag{12}$$

3.2.2 Detection association

The maximum matching of problem (8) can be solved using the Hungarian algorithm. The final aim is to find the best match for each trajectory from the candidates. Given the detections D^i and the set of trajectories $\mathcal{T}^{i-1}$, the association matrix used by the Hungarian algorithm can be computed as

$$\boldsymbol{S} = \begin{bmatrix} g_1^1 & g_1^2 & \cdots & g_1^r \\ g_2^1 & g_2^2 & \cdots & g_2^r \\ \vdots & \vdots & & \vdots \\ g_s^1 & g_s^2 & \cdots & g_s^r \end{bmatrix}, \tag{13}$$

where g_j^p ($j=1,\ 2,\ \ldots,\ s$; $p=1,\ 2,\ ,\ldots,\ r$) is the association score, calculated by $\alpha_1\cdot\Psi_1(\boldsymbol{e}_j^{i-1},\ \boldsymbol{d}_p^i)+\alpha_2\cdot\Psi_2(\boldsymbol{e}_j^{i-1},\ \boldsymbol{d}_p^i)+\alpha_3\cdot\Psi_3(\boldsymbol{e}_j^{i-1},\ \boldsymbol{d}_p^i)+\alpha_4\cdot\Psi_4(\boldsymbol{e}_j^{i-1},\ \boldsymbol{d}_p^i)$, $\boldsymbol{d}_p^i$ and $\boldsymbol{e}_j^{i-1}$ are the detection from D_i and the point from ξ^{i-1} respectively, $r=|D^i|$, and $s=|\xi^{i-1}|$ ($|\cdot|$ represents the number of elements in the set). Through greedily searching by the Hungarian algorithm, the optimal correspondences between detections and trajectories are obtained, and then the end points of the trajectories are updated accordingly. The overall association algorithm is detailed in Algorithm 2, and some false detections can be inhibited in the association process.

We can find that strong noise or clutter can also result in long false trajectories in the output

Algorithm 2 Detection association

Input: $\{D^i\}_{i=l+m,\ l+m+1,\ \ldots,\ L-l}$
Output: $\mathcal{T}^{L-l}$
1: Initialize the set of trajectories $\mathcal{T}^{i-1}$ with D^{i-1}
2: **for** each combination ($\mathcal{T}^{i-1}$, D^i) **do**
3: Compute the association matrix $\boldsymbol{S}$
4: Obtain the best matches between $\mathcal{T}^{i-1}$ and D^i using the Hungarian algorithm
5: Append the assigned trajectories $\{\boldsymbol{T}_j^{i-1}\}$ with the matched detections $\{\boldsymbol{d}_p^i\}$ to form the set $\mathcal{T}^i$
6: Delete trajectories that are not assigned for several times
7: Create new trajectories $\{\boldsymbol{T}_{\text{new}}\}$ for unmatched detections
8: Append $\{\boldsymbol{T}_{\text{new}}\}$ to the set $\mathcal{T}^i$
9: $i-1\leftarrow i$
10: **end for**

of Algorithm 2, and the false trajectories are diverse from the true trajectories produced by real targets. According to this point, for the set of trajectories $\mathcal{T}^{L-l}=\{\boldsymbol{T}_1^{L-l},\ \boldsymbol{T}_2^{L-l},\ \ldots,\ \boldsymbol{T}_j^{L-l},\ \ldots\}$, we define a displacement gain measure Δ_j to distinguish between the true and false trajectories:

$$\Delta_j = \sum_n |x_n - x_{n-1}| + |y_n - y_{n-1}|, \tag{14}$$

where (x_n, y_n) and (x_{n-1}, y_{n-1}) denote the coordinates of points from $\boldsymbol{T}_j^{L-l}$. Let θ be the threshold of displacement gains. If the gain Δ_j is larger than θ, the trajectory can be verified as a true trajectory; otherwise, the trajectory can be considered as a false trajectory, and should be discarded accordingly.

4 Experimental results

We used four infrared sequences to test the proposed detection method. The first three sequences (NPA, J2A, and NA23A) were taken by the American Air Force Rome Laboratory, and each of them contains 95 frames with a size of 244×320 pixels. The last sequence (IRPL) was taken from the OTCBVS benchmark (Miezianko, 2006), consisting of 100 frames with a size of 240×320 pixels. To evaluate the performance of the proposed detection method, we chose five methods, i.e., nonlinear adaptive filtering (NAF) (Liu *et al.*, 2015), spatial-temporal local contrast (STLC) (Deng *et al.*, 2016), local contrast (LC) (Chen *et al.*, 2014), temporal contrast (TC) (Kim *et al.*, 2014), and patch-image (PI) (Gao *et al.*, 2013), as the baseline methods.

Fig. 2 shows the detection results of the proposed temporal target detection method. Four small segments were selected respectively from IRPL (a landing target with clutter of skyline and plants), NPA (flying targets on a background of wispy clouds), NA23A (a flying target on a background of small bright clouds), and J2A (flying targets on a background of fluffy clouds) for test data. We can find that the results from Figs. 2b and 2c have wildly different characteristics. Cloud edges and random noise are the main clutter in Fig. 2b, but there are very few cloud edges and random noise in Fig. 2c. So, the cross-product operation can be used to suppress clutter and noise to reduce false alarms. We can see from Fig. 2d that most clutter and strong noise is inhibited while detecting targets, but some small errors still exist. Therefore, the association strategy based on target movement characteristics is required to validate false alarms and to improve the detection performance.

To evaluate the detection performance from a subjective perspective, Fig. 3 gives the results of our proposed method and five baseline methods (NAF, STLC, TC, LC, and PI) on the four IR sequences. As shown in Figs. 3b–3d, the NAF, STLC, and TC methods can detect most target pixels through temporal filtering, but the false alarms (such as the cloud edges and noise in the detection results) still exist and could degrade the detection performance. As shown in Figs. 3e and 3f, the LC and PI methods can detect most target pixels on the IRPL sequence, but cannot work well on sequences NPA, NA23A,

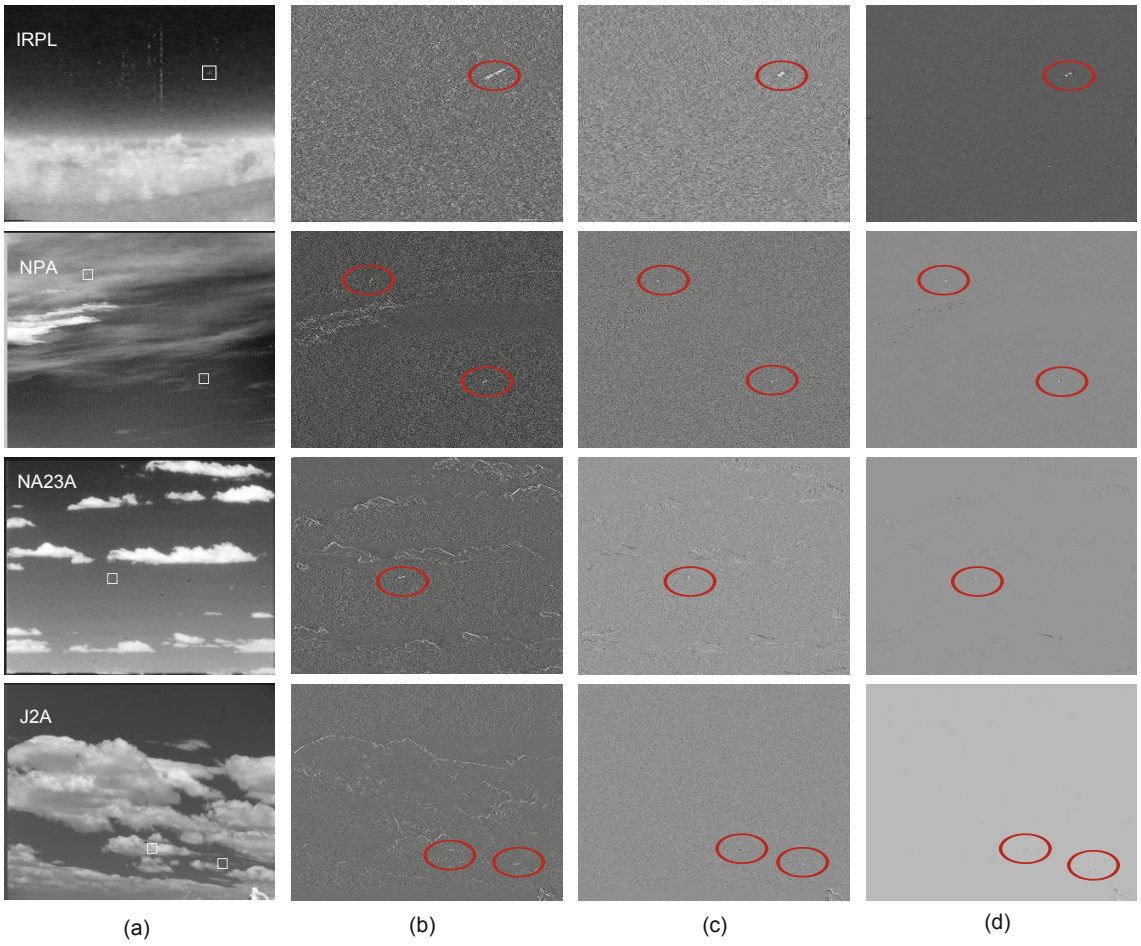

Fig. 2 Detection results of the proposed temporal target detection method: (a) original images of four sequences (IRPL, NPA, NA23A, and J2A); (b) detection results of temporal max filtering, temporal target fusion, and SVD reconstruction; (c) detection results of temporal median filtering, temporal target fusion, and SVD reconstruction; (d) cross-product filtering of the detection results from (b) and (c)

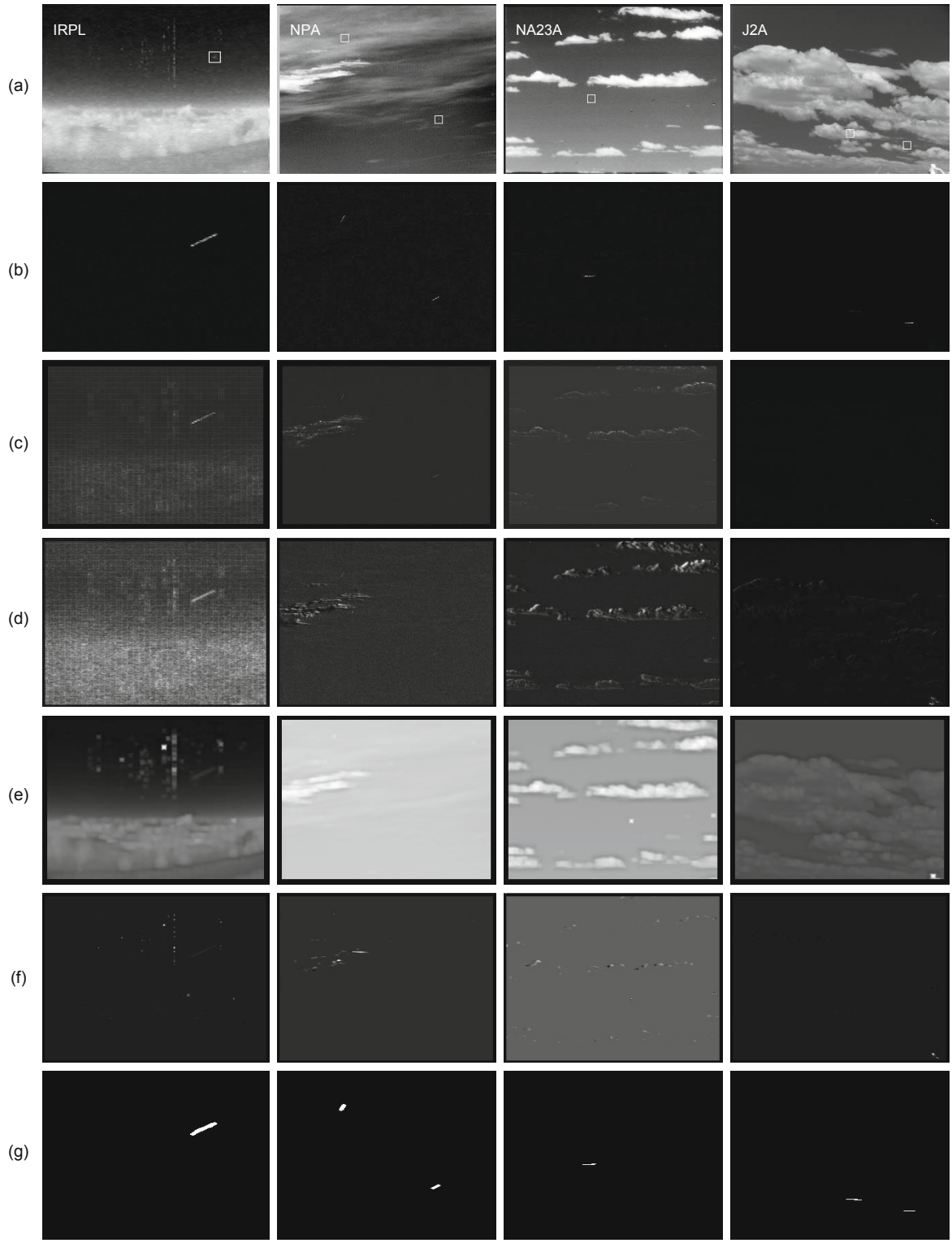

Fig. 3 Comparison of detection results of the proposed method and five baseline methods on four sequences (IRPL, NPA, NA23A, and J2A): (a) original images; (b) nonlinear adaptive filtering (NAF) method; (c) spatial-temporal local contrast (STLC) method; (d) temporal contrast (TC) method; (e) local contrast (LC) method; (f) patch-image (PI) method; (g) our proposed method

and J2A with bright cloud clutter. This is because the NAF, STLC, and TC methods can make full use of the target and clutter information of consecutive frames, but the LC and PI methods use only the spatial target information, and do not consider the movement characteristics of the slowly moving targets. As shown in Figs. 3b–3d, the temporal cloud edges and noise are easily mistaken for targets. Although the NAF method has better detection performance in comparison with the STLC and TC methods, when further suppressing clutter by normal threshold segmentation, some target pixels can be considered as false alarms, and be discarded. Therefore, multiplication of two temporal filters and association of moving targets are necessary to improve the temporal detection result. As shown in Fig. 3g, the proposed method can not only suppress the cloud clutter and strong noise very well, but also detect the target pixels more accurately than the baseline methods.

The detection performance can also be evaluated by varying the detection thresholds for respective methods. For any method, each threshold corresponds to a pair (P_a, P_d). All the pairs are then plotted to form a receiver operating characteristic (ROC) curve by defining P_d as a function of P_a. The probabilities of correct detection (P_d) and false-alarm (P_a) are respectively defined as follows (Gao *et al.*, 2013; Bae, 2014):

$$P_\mathrm{d} = \frac{\text{number of correctly detected pixels}}{\text{number of actual target pixels}}, \quad (15)$$

$$P_\mathrm{a} = \frac{\text{number of incorrectly detected pixels}}{\text{number of total false pixels}}. \quad (16)$$

Therefore, the ROC curve can be used to estimate the rate of correctly detected target pixels and incorrectly detected pixels in the result. Note that a higher score of P_d indicates a better detection result, and a higher score of P_a specifies a worse performance.

The probabilities of correct detection and false-alarm of the proposed method are dependent on the threshold of displacement gains, but the correct detection rates and the false-alarm rates of the five baseline methods rely on their respective segmentation thresholds. Fig. 4 shows the ROC curves of the proposed method and the five baseline methods computed using the four IR sequences. The mean values of P_d and P_a for the six methods are listed in Table 1. Although the NAF method has a higher average detection rate than the other methods on the IRPL sequence, it has a higher false-alarm rate than the proposed method. Compared with the baseline methods, the proposed method has the highest average detection rates and lowest average false-alarm rates on sequences NPA, NA23A, and J2A.

Figs. 4b–4d also show that for the IR backgrounds with bright clouds (NPA, NA23A, and J2A), the temporal filtering methods (such as the proposed method, NAF, STLC, and TC) have better performance than the spatial filtering methods (such as LC and PI), because the latter cannot effectively filter out the strong cloud edges while simultaneously ensuring high detection rates. For the IR background with strong noise (IRPL), the spatial methods (such as the PI method) may achieve better detection results than other temporal methods (such as the STLC method), by choosing appropriate parameters (Fig. 4a). However, in general, the temporal methods are superior to the spatial methods when dealing with the IR sequences with slowly moving complex backgrounds. This is because the former can model the target and clutter information of consecutive frames. To better use the merits of temporal filtering methods and to improve the detection

Table 1 Mean values of P_d and P_a for the proposed method and five baseline methods

Method	Mean value of P_d				Mean value of P_a			
	IRPL	NPA	NA23A	J2A	IRPL	NPA	NA23A	J2A
NAF	0.9480	1.0000	0.9230	0.8390	0.0740	0.1230	0.1050	0.0220
STLC	0.3620	0.5030	0.2310	0.4760	0.1040	0.0810	0.1250	0.0700
TC	0.2160	0.2490	0.0960	0.2490	0.1690	0.1080	0.0980	0.0920
LC	0.0920	0.0770	0.0770	0.1100	0.1520	0.0720	0.0890	0.0950
PI	0.2710	0.1070	0.0770	0.1060	0.0740	0.0750	0.0750	0.0740
Proposed	0.9300	1.0000	1.0000	1.0000	0.0010	0.0010	0.0020	0.0003

NAF: nonlinear adaptive filtering method; STLC: spatial-temporal local contrast method; TC: temporal contrast method; LC: local contrast method; PI: patch-image method

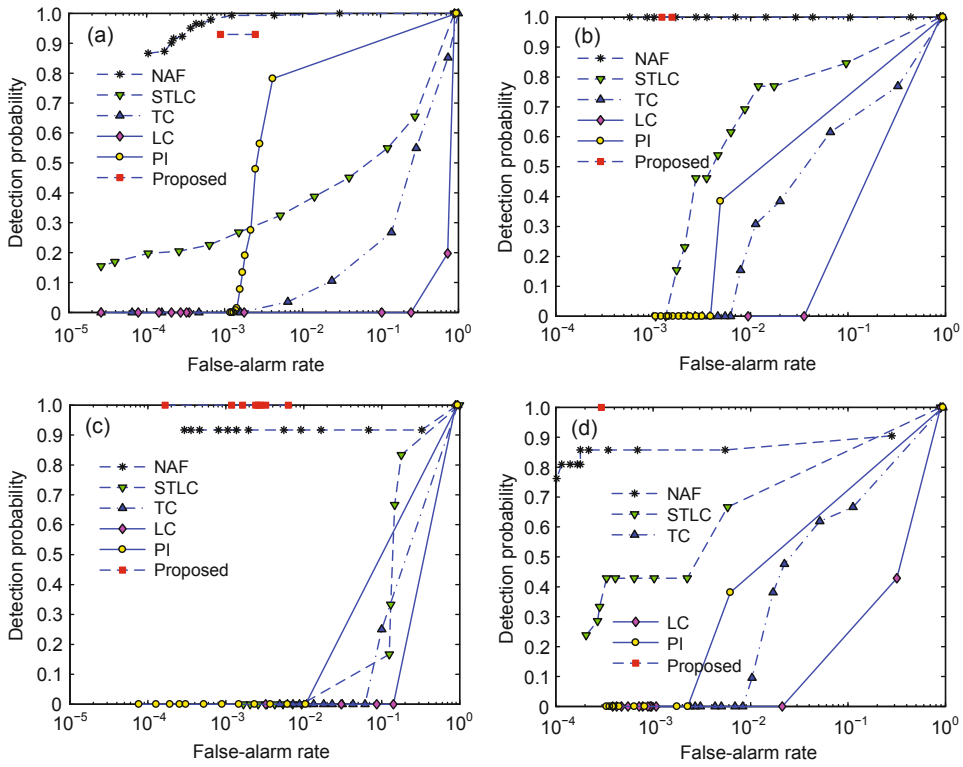

Fig. 4 ROC curves of the proposed method and five baseline methods obtained from sequences IRPL (a), NPA (b), NA23A (c), and J2A (d)

performance, the proposed method suppresses heavy noise and clutter using two complementary temporal filters (see Fig. 2), and moreover inhibits false alarms in terms of the association strategy and movement characteristics of moving targets. Although some baseline methods, such as the NAF method, can detect the target trajectory effectively, the number of targets in an IR sequence is still not known. Different from the baseline methods, based on the target trajectories obtained, the proposed method can determine the target number and suppress the false alarms further. We can find that for the IRPL sequence, the minimum P_{d} (estimating the precision of a target trajectory) of the proposed method is only 0.93, but it is enough to guarantee the existence of a target trajectory.

Table 2 shows the running time of the proposed method and five baseline methods on the four IR sequences. Although the STLC and TC methods are superior to the other methods from the view of computation time, they have inferior detection performance. The running time of the proposed method can occupy the third place in these six algorithms. It is only slightly better than that of the NAF method, but overall the proposed method can achieve better detection performance than the other methods (see Table 2 and Figs. 3 and 4).

Table 2 Running time of the proposed method and five baseline methods on the four sequences

Method	Running time (s)			
	IRPL	NPA	NA23A	J2A
NAF	53.14	55.71	53.22	52.05
STLC	18.31	23.50	22.09	22.42
TC	10.10	9.71	9.56	9.75
LC	89.51	85.23	85.49	85.36
PI	1020.99	939.97	911.44	890.65
Proposed	51.74	50.19	52.60	50.83

NAF: nonlinear adaptive filtering method; STLC: spatial-temporal local contrast method; TC: temporal contrast; LC: local contrast method; PI: patch-image method

5 Conclusions

It is rather difficult to detect the slowly moving targets buried in complex backgrounds due to the

lack of rich features, such as containing only a few pixels, shapeless edge, and a low signal-to-clutter ratio. In this paper, an infrared target detection algorithm based on temporal filtering and target association has been proposed. Bright cloud clutter and strong noise can be suppressed based on the proposed temporal target detection model, and false alarms can be further reduced using the target association method. Real data experiments show that the proposed method has better detection performance in comparison with several baseline methods. Moreover, the temporal filtering methods are superior to the spatial filtering methods when detecting small targets immersed in bright cloud backgrounds.

References

Bae, T., 2014. Spatial and temporal bilateral filter for infrared small target enhancement. *Infrared Phys. Technol.*, **63**:42-53. http://dx.doi.org/10.1016/j.infrared.2013.12.007

Chen, C.L., Li, H., Wei, Y., *et al.*, 2014. A local contrast method for small infrared target detection. *IEEE Trans. Geosci. Remote Sens.*, **52**(1):574-581. http://dx.doi.org/10.1109/TGRS.2013.2242477

Chen, Z., Wang, X., Sun, Z., *et al.*, 2016. Motion saliency detection using a temporal Fourier transform. *Opt. Laser Technol.*, **80**:1-15. http://dx.doi.org/10.1016/j.optlastec.2015.12.013

Comaniciu, D., Ramesh, V., Meer, P., 2000. Real-time tracking of non-rigid objects using mean shift. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, p.142-149. http://dx.doi.org/10.1109/CVPR.2000.854761

Deng, L., Zhu, H., Tao, C., *et al.*, 2016. Infrared moving point target detection based on spatial-temporal local contrast filter. *Infrared Phys. Technol.*, **76**:168-173. http://dx.doi.org/10.1016/j.infrared.2016.02.010

Deshpande, S.D., Meng, H.E., Venkateswarlu, R., *et al.*, 1999. Max-mean and max-median filters for detection of small targets. Proc. SPIE, p.74-83. http://dx.doi.org/10.1117/12.364049

Dong, X., Huang, X., Zheng, Y., *et al.*, 2014. Infrared dim and small target detecting and tracking method inspired by human visual system. *Infrared Phys. Technol.*, **62**:100-109. http://dx.doi.org/10.1016/j.infrared.2013.11.007

Gao, C., Meng, D., Yang, Y., *et al.*, 2013. Infrared patch-image model for small target detection in a single image. *IEEE Trans. Image Process.*, **22**(12):4996-5009. http://dx.doi.org/10.1109/TIP.2013.2281420

Gao, J., Wen, C., Liu, M., 2015. Low-speed small target detection based on SVD and superposition. *J. Shanghai Jiao Tong Univ.*, **49**(6):876-883 (in Chinese). http://dx.doi.org/10.16183/j.cnki.jsjtu.2015.06.023

Kim, S., Sun, S.G., Kim, K.T., 2014. Highly efficient supersonic small infrared target detection using temporal contrast filter. *Electron. Lett.*, **50**(2):81-83. http://dx.doi.org/10.1049/el.2013.2109

Li, Y., Li, P., Shen, Q., 2014. Real-time infrared target tracking based on ℓ_1 minimization and compressive features. *Appl. Opt.*, **53**(28):6518-6526. http://dx.doi.org/10.1364/AO.53.006518

Liu, D., Li, Z., Wang, X., *et al.*, 2015. Moving target detection by nonlinear adaptive filtering on temporal profiles in infrared image sequences. *Infrared Phys. Technol.*, **73**:41-48. http://dx.doi.org/10.1016/j.infrared.2015.09.003

Liu, R., Li, X., Han, L., *et al.*, 2013. Track infrared point targets based on projection coefficient templates and non-linear correlation combined with Kalman prediction. *Infrared Phys. Technol.*, **57**:68-75. http://dx.doi.org/10.1016/j.infrared.2012.12.011

Miezianko, R., 2006. IEEE OTCBVS WS Series Bench: Terravic Research Infrared Database. Available from http://vcipl-okstate.org/pbvs/bench/Data/05/download.html.

Silverman, J., Mooney, J.M., Caefer, C.E., 1996. Temporal filters for tracking weak slow point targets in evolving cloud clutter. *Infrared Phys. Technol.*, **37**(6):695-710. http://dx.doi.org/10.1016/S1350-4495(96)00003-5

Taj, M., Maggio, E., Cavallaro, A., 2006. Multi-feature graph-based object tracking. Proc. 1st Int. Evaluation Workshop on Classification of Events, Activities and Relationships, p.190-199. http://dx.doi.org/10.1007/978-3-540-69568-4_15

Wang, Z., Tian, J., Liu, J., *et al.*, 2006. Small infrared target fusion detection based on support vector machines in the wavelet domain. *Opt. Eng.*, **45**(7):076401. http://dx.doi.org/10.1117/1.2218864

Wang, Z., Ma, Y., Wang, L., 2013. Assessment of threat degree for LSS target in air defense operation. *Shipboard Electron. Countermeas.*, **36**(6):103-105 (in Chinese).

Yan, X., Wu, X., Kakadiaris, I.A., *et al.*, 2012. To track or to detect? An ensemble framework for optimal selection. Proc. 12th European Conf. on Computer Vision, p.594-607. http://dx.doi.org/10.1007/978-3-642-33715-4_43

Yang, Y., Wu, J., Zheng, W., 2012. Trajectory tracking for an autonomous airship using fuzzy adaptive sliding mode control. *J. Zhejiang Univ.-Sci. C (Comput. & Electron.)*, **13**(7):534-543. http://dx.doi.org/10.1631/jzus.C1100371

Zhang, F., Li, C., Shi, L., 2005. Detecting and tracking dim moving point target in IR image sequence. *Infrared Phys. Technol.*, **46**(4):323-328. http://dx.doi.org/10.1016/j.infrared.2004.06.001

Zhang, J., Guo, H., 2012. Net cast interception system research aimed at low small slow target. *Comput. Eng. Des.*, **33**(7):2874-2878 (in Chinese).

Zhang, J., Li, Q., Cheng, N., *et al.*, 2013. Nonlinear path-following method for fixed-wing unmanned aerial vehicles. *J. Zhejiang Univ.-Sci. C (Comput. & Electron.)*, **14**(2):125-132. http://dx.doi.org/10.1631/jzus.C1200195

Zhang, Y., Xin, Y., Zhang, C., 2010. An algorithm based on temporal and spatial filters for infrared weak slow moving point target detection. *Acta Photon. Sin.*, **39**(11):2049-2054 (in Chinese). http://dx.doi.org/10.3788/gzxb20103911.2049