Flying Small Target Detection for Anti-UAV Based on a Gaussian Mixture Model in a Compressive Sensing Domain

Chuanyun Wang 1,*, Tian Wang 2, Ershen Wang 3, Enyan Sun 1 and Zhen Luo 1

1 School of Computer Science, Shenyang Aerospace University, Shenyang 110136, China; sunenyan@sau.edu.cn (E.S.); luozhen@sau.edu.cn (Z.L.)
2 School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China; wangtian@buaa.edu.cn
3 School of Electronic and Information Engineering, Shenyang Aerospace University, Shenyang 110136, China; wes2016@sau.edu.cn
* Correspondence: wangcy0301@sau.edu.cn; Tel.: +86-1581-136-6220

Received: 5 April 2019; Accepted: 6 May 2019; Published: 10 May 2019

Abstract: Addressing the problems of visual surveillance for anti-UAV, a new flying small target detection method is proposed based on Gaussian mixture background modeling in a compressive sensing domain and low-rank and sparse matrix decomposition of local image. First of all, images captured by stationary visual sensors are broken into patches and the candidate patches which perhaps contain targets are identified by using a Gaussian mixture background model in a compressive sensing domain. Subsequently, the candidate patches within a finite time period are separated into background images and target images by low-rank and sparse matrix decomposition. Finally, flying small target detection is achieved over separated target images by threshold segmentation. The experiment results using visible and infrared image sequences of flying UAV demonstrate that the proposed methods have effective detection performance and outperform the baseline methods in precision and recall evaluation.

Keywords: flying small target detection; anti-UAV; gaussian mixture model; compressive sensing; low-rank and sparse matrix decomposition

1. Introduction

In recent years, the Unmanned Aerial Vehicles (UAV) industry has been booming at a rate and scale and has gained emerging interest in both civilian and military applications. Especially, the consumer micro-UAV, such as drones, is growing exponentially and joins in the airspace setting for commercial and individual needs, such as aviation videography, courier services, traffic surveillance and land surveying [1]. However, as the rapid spread of micro-UAV, many new security threats and social problems, including disturbing aviation safety of commercial aircraft, breaching no-fly security in sensitive areas and invading public privacy, frequently occur and appear to be continually increasing [2].

In spite of the laws and regulations being enforced and no-fly zone control being built in micro-UAV to restrain operators’ behavior, there still exists potentially dangers due to the unauthorized flying with spontaneous. As a consequence, there is a critical need for an anti-UAV system to detect a micro-UAV that enters a specific sensitive airspace in time. The traditional technology of anti-UAV is radar which works by means of active illumination with electromagnetic waves [3]. In the low-attitude airspace context, the radar signals of targets immerse in complex background clutter, e.g., trees, buildings [4,5]. Moreover, potentially dangerous targets should be detected when they are still far
away, which means they may still be very small. Thereupon, the cost performance of expensive radar for micro-UAV detection is greatly reduced.

Fortunately, vision sensors are low-cost, passive and robust to jamming; hence, motivating their use for micro-UAV detection, even though they have a limited range compared to radars. As a matter of fact, there still exists many challenges to detect a micro-UAV with a visual method. This is because it not only occupies a small portion of the field of view from a far distance and flies against complex backgrounds under various weather and lighting conditions, but also appears changing size and diverse shapes in fully 3D dimensional environment.

Over the past few years, the problems of anti-UAV have been attracting more and more attention, but only a few studies of visual detection are studied. A non-technical way of anti-UAV was gathering information about potentially illegal or dangerous activity of UAV via crowd sourcing from their mobile phone by submitting a short online DroneALERT incident report [6]. The drone surveillance technologies and the existing anti-drone systems are comprehensively reviewed [2], and an anti-drone system called ADS-ZJU is developed by utilizing audio, video and RF technologies [7]. Sosnowski et al. presented a concept of short-range air defence systems to protect against drone threats with various optoelectronic sensors [8]. Inspired by different optical properties possessed by small drones with the sky as a background, a combination of infrared polarization filters was suggested to identify drones [9]. In addition, multiple local features of aircraft was used to compose a feature clustering criterion for vision based aircraft tracking [3]. Furthermore, some researchers have attempted to detect and track flying small UAVs from a single moving camera. Li et al. classified candidate targets by using their spatio-temporal traits on the basis of background motions estimation and distinctive points identification, and then a Kalman filter was used to boost the tracking performance [10]. Rozantsev et al. achieved object-centric motion stabilization of image patches by a regression-based approach, and then detected flying objects by combining both appearance and motion cues as spatio-temporal image cubes [11,12]. In addition to these, the research of vision-based sense-and-avoid for micro-UAV could provide a reference [13,14].

In this study, multiple visual sensors are distributed around the surveillance area, and no-blind area coverage for surveillance is realized in a cooperative manner, so as to build an intelligent visual surveillance system for discovering the approaching UAV in time [15,16]. However, due to the optical properties and image resolution of visual sensors, it is limited to detect a target as far as possible. In that case, the size of a target is often very small and with only a few pixels because of a long imaging distance, as well as the energy of the target received by an imaging sensor is very low and with dim spotlike features because of atmospheric effects [17–19]. Therefore, an efficient flying small target detection algorithm is crucial to give a warning as early as possible.

In this paper, we concentrate on flying small target detection for anti-UAV from a stationary camera. Making use of the advantages of compressive sensing which can efficiently compress a higher-dimensional data to lower-dimensional one and perfectly reconstruct from a small number of linear measurements with high probability, the image background modeling for flying small target detection is no longer over each pixel but over each linear measurement. Subsequently, the images in sequence are converted into patches to construct a background model for better enforcing the image neighboring correlation, and then the candidate patches which perhaps contain targets that are identified by using background models of each patch. Moreover, with the characteristics of background scene slight change and target relative salient movement within a finite time period, a matrix composed of images in sequence is separated into background images and target images by low-rank and sparse matrix decomposition. As a Consequence, flying small target detection for anti-UAV is achieved over a separated target image.

The remainder of this paper is organized as follows. Firstly, a new Gaussian mixture model in a compressive sensing domain is designed in Section 2. Section 3 presents the approach to flying small target detection composed by candidate patches identification, local foreground-background images separation and comprehensive scheme of the proposed method. Section 4 contains experiments to
demonstrate the effectiveness of the proposed method and compare its performance with baseline methods. Finally, conclusions are drawn in Section 5.

In order to clearly depict this work, the major notations used throughout the rest of this paper are briefly summarized in Table 1.

| Notations | Descriptions |
|-----------|--------------|
| s         | the number of Gaussian distribution |
| w_i       | the weight of the i_th Gaussian distribution |
| µ_i       | the mean of the i_th Gaussian distribution |
| σ^2_i     | the variance of the i_th Gaussian distribution |
| x         | a vector stacked by an image with n pixels |
| y         | a compressive sensing measurement vector with m dimensions |
| Ψ         | a sparse basis matrix composed by q n-dim basis vectors |
| Φ         | a measurement matrix composed by m n-dim Gaussian distribution |
| Θ         | the weighting coefficients vector of x in the Ψ domain |
| ξ         | a constant for Gaussian distribution identification |
| α         | the learning rate of weight of the Gaussian distribution |
| β         | the learning rate of mean and variance of the Gaussian distribution |
| τ         | the value of the Gaussian distribution matched or not |
| J_k       | the value of the k_th linear measurement matched or not |
| δ         | a threshold for identifying image patch maintain target or not |
| D         | a matrix composed by r rows and c columns block matrices |
| G         | a matrix composed by a group of vectorized consecutive patches |
| L         | the low-rank matrix decomposed from matrix G via RPCA |
| S         | the sparse matrix decomposed from matrix G via RPCA |
| R         | the rank of the low-rank matrix L |
| K         | the sparse degree of the sparse matrix S |
| λ         | a weight parameter for tradeoff the matrices L and S |
| u_i       | the i_th column vector of matrix U after SVD |
| v_i       | the i_th column vector of matrix V after SVD |
| Ω_L       | a subset of the entire set of singular vectors O |
| Ω_S       | a subset of the entire set of position support terms Q |

2. Gaussian Mixture Modeling in a Compressive Sensing Domain

Traditionally, background modeling has been attracting significant attention in recent years for target detection in the foreground. To tackle this problem, the Gaussian mixture model (GMM) is proposed as a statistical approach by training multiple models of probability density functions over each pixel [20,21]. If a variable x is Gaussian distribution with mean µ and variance σ^2, its probability density function (PDF) is defined as

\[ N(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \]  

However, the gaussian distribution is a single modal distribution, which can not provide a good approximation for the data distribution of multimodal states. Fortunately, according to previous literatures, GMM is considered to fit any probability distribution of a variable, that is to say, any variable can be regarded as with probability of multiple specific gaussian distributions [22]. Consequently, the GMM is composed by multiple Gaussian distributions, and its PDF is written as the linear sum of PDFs of the Gaussian distributions

\[ p(x) = \sum_{i=1}^{s} p(i) p(x|i) = \sum_{i=1}^{s} w_i N(x; \mu_i, \sigma_i^2) \]
where \( s \) is the number of Gaussian distribution, \( w_i, \mu_i \) and \( \sigma_i^2 \) are the weight, mean and variance of the \( ith \) Gaussian distribution, respectively. In general, \( s \) is set to \( 3 \sim 5 \), and the Gaussian distributions are independent to each other [23,24]. In this study, the flying target with salient movement is the foreground, while the background scene changes are relatively slight. Thereupon, a Gaussian distribution will catch the pattern of follow-up compressive sensing measurements.

The pixel-based background modeling algorithms establish models for each pixel of an image and classify a new pixel as foreground if its PDF value is below a threshold [25]. Therefore, background modeling can be performed on these linear measurements, as they are a low-dimensional compressive representations of the given signal.

An image is stacked to form a vector \( x = [x_1, x_2, \ldots, x_n]^T \in \mathbb{R}^n \). The linear projection of the image vector on the measurement matrix \( \Phi \in \mathbb{R}^{m \times n} (m \ll n) \) is defined as

\[
y = \Phi x
\]

where \( y = [y_1, y_2, \ldots, y_n]^T \in \mathbb{R}^m \) represents the low dimensional CS measurement vector.

However, image itself may not be sparse but can be sparsely represented in some transform domain, such as DCT, Fourier, wavelet and over-complete dictionary. Assume that \( x \in \mathbb{R}^m \) can be sparsely represented in terms of a basis of vectors \( \{ \psi_i \}_{i=1}^{q} \) with \( \psi_i \in \mathbb{R}^n \), then forming the sparse basis matrix \( \Psi = [\psi_1, \psi_2, \ldots, \psi_q] \in \mathbb{R}^{n \times q} \) by stacking the vectors as columns, so \( x \) can express as

\[
x = \sum_{i=1}^{q} \psi_i \theta_i = \Psi \Theta
\]

where \( \Theta = [\theta_1, \theta_2, \ldots, \theta_q]^T \in \mathbb{R}^q \) is the weighting coefficients vector of \( x \) in the \( \Psi \) domain [28].

Based on the above discussion, compressive sensing of an image can be described as linear projection process [29]

\[
y = \Phi \Psi \Theta = A \Theta
\]

where \( A := \Phi \Psi \) is referred to as sensing matrix. In order for the well-conditioned reconstruction of a \( K \)-sparse signal \( x \) in the specific transform domain, the sensing matrix \( A \in \mathbb{R}^{m \times q} \) should satisfies the restricted isometry property (RIP) with order \( 2K \) [30].

**Definition 1.** A matrix \( A \) satisfies the restricted isometry property (RIP) of order \( K \) if there exists a \( \delta_K \in (0,1) \) such that

\[
(1 - \delta_K) \| \Theta \|_2^2 \leq \| A \Theta \|_2^2 \leq (1 + \delta_K) \| \Theta \|_2^2,
\]

for all \( \Theta \in \{ \Theta : \| \Theta \|_0 \leq K \} \), where \( \| \Theta \|_0 := \sharp \{ i | \theta_i \neq 0 \} \) is the zero pseudo-norm [31].

Unfortunately, the task of holding that the sensing matrix satisfies the RIP is an NP-hard problem, but an alternative approach is to make sure that the measurement matrix \( \Phi \) is incoherent with the sparse basis matrix \( \Psi \) [29]. The greater incoherence of the measurement/sparsity pair \((\Phi, \Psi)\), the smaller number of the linear measurements needed. In a Gaussian random matrix, the elements are independent and identically distributed (i.i.d) random variables from a zero-mean, \( 1/n \)-variance Gaussian density. Then, a Gaussian random matrix is incoherent with most of orthogonal sparse
basis matrix, and is chosen as measurement matrix of compressive sensing to collect information of random projections

$$\Phi = [\phi_1, \phi_2, \cdots, \phi_m]^T$$  \hspace{1cm} (6)

where $\phi_i \in \mathbb{R}^n$, and $\phi_i$ is Gaussian distribution with $n$-dimension $\mathcal{N}(0, I)$.

With these considerations in mind, we replace the pixels with a small set of CS measurements to establish a background model for flying small target detection [32], employing the low dimensional CS measurement vector $y = [y_1, y_2, \cdots, y_m]^T$ instead of the high dimensional image vector $x = [x_1, x_2, \cdots, x_n]^T$ to establish Gaussian mixture model by Equation (2), and this process is so-called Gaussian mixture modeling in a compressive sensing domain (GMM_CS) [33].

When a new image is captured at time $t$ during model training, a series of linear projections with Gaussian random matrix is operated over the image vector by Equation (3). Subsequently, each element in CS measurement vector $y$ is identified with previous Gaussian mixture model by

$$|y_k^t - \mu_{k,i}^{t-1}| < \xi \sigma_{k,i}^{t-1} \quad (i = 1, 2, ..., s)$$  \hspace{1cm} (7)

where $\xi$ is a constant. If there exists a matched Gaussian distribution, the corresponding mean and variance should be updated by utilizing current CS measurement

$$\mu_k^t = (1 - \beta)\mu_k^{t-1} + \beta y_k^t$$  \hspace{1cm} (8)

$$(\sigma_k^t)^2 = (1 - \beta)(\sigma_k^{t-1})^2 + \beta(y_k^t - \mu_k^t)^2$$  \hspace{1cm} (9)

where $\beta \in (0, 1)$ represents the learning rate of mean and variance of the Gaussian distribution. Then, the weights of $s$ Gaussian distributions are updated simultaneously

$$w_{k,i}^t = (1 - (-1)^T \alpha)w_{k,i}^{t-1}$$  \hspace{1cm} (10)

where $\alpha \in (0, 1)$ denotes the learning rate of weight of the Gaussian distribution, $\tau \in \{0, 1\}$ will set to 1 for the matched Gaussian distribution and to 0 for mismatched one. If there no matched Gaussian distribution, the mean and variance of the Gaussian distribution with minimum weight will be updated, and then the weights of $s$ Gaussian distributions are normalized and let $\sum_{i=1}^{s} w_{k,i} = 1$.

3. Approach to Flying Small Target Detection by Local Image Separation

3.1. Candidate Patches Identification Based on GMM_CS

From a general viewpoint, images captured by imaging systems for fly small target detection have large sizes, thus, when we treat a two-dimensional image by vectorizing it into a one-dimensional vector, the corresponding vector is with high dimensions [34]. As a consequence, there are two negative influences. On one hand, a fairly large measurement matrix is needed for CS which makes the storage and computations of a Gaussian ensemble very difficult [35]. On the other hand, the linear projection process over the high dimensional vector of whole image reckons without the image neighboring correlation.

In addressing the problems above, we break the whole image into overlapping or non-overlapping patches by patch transform [36]. Assume a new image is represented by a matrix $D$, the non-overlapping patch transform will convert the matrix into several block matrices with $r$ rows and $c$ columns

$$D = \begin{bmatrix}
D_{1,1} & D_{1,2} & \cdots & D_{1,c} \\
D_{2,1} & D_{2,2} & \cdots & D_{2,c} \\
\vdots & \vdots & \ddots & \vdots \\
D_{r,1} & D_{r,2} & \cdots & D_{r,c}
\end{bmatrix}$$  \hspace{1cm} (11)
where $D_{i,j}$ denotes the $i^{th}$ row, $j^{th}$ column image patch represented by a block matrix. For simplicity, each train image and test image are divided into non-overlapping patches for experiment in this study. In spite of the small size of the image patch leading to low computational complexity for one image patch, more image patches which perhaps maintain a flying small target, and subsequently, high computational complexity will be induced in the following process of target detection. On account of these considerations, we designate the size of image patch to refer to the roughly maximum size of flying target and division into integer patches.

Subsequently, the background modeling and target detection procedures will be applied over image patches. The background modeling is to establish a Gaussian mixture model for each linear measurement of image patches generated from a group of training images. After that, the obtained background model based on GMM_CS can be employed to identify the candidate patches which perhaps maintain a flying small target. Thereinto, the identification of an image patch maintains a flying small target, operates in a compressive sensing domain, and includes two steps. The first step is the comparison between each linear measurement of input image patch with the obtained background model by Equation (7); the second step is to compute the statistical ratio of matched linear measurements to the total linear measurements, which is expressed as

$$\delta = \frac{1}{m} \sum_{k=1}^{m} J_k$$

(12)

where $\delta \in (0, 1]$ is regard as a threshold for identifying current image patch is maintain a flying small target or not, $m$ is the total linear measurements of current image patch, and $J$ is defined as

$$J_k = \begin{cases} 1, & \text{if } |y_{k,i} - \mu_{k,i}| < \xi \sigma_{k,i} \\ 0, & \text{otherwise} \end{cases} \quad (i = 1, 2, ..., s)$$

(13)

Based on the aforementioned discussion, a demonstration of candidate patches identification based on GMM_CS is depicted in Figure 1.

**Figure 1.** A demonstration of candidate patches identification based on GMM_CS.

### 3.2. Local Target-Background Patches Separation

Along with continuously captured scene images of the imaging system, a group of consecutive candidate patches extracted form several adjacent images is designated and vectorized as columns to compose a data matrix $G \in \mathbb{R}^{d \times l}$, where $d$ is the total number of pixels in each candidate patch, and $l$ is the number of adjacent images. Making use of the characteristics of background scene slight change and target relative salient movement within a finite time period, the data matrix $G$ will be decomposed to a low-rank matrix $L$ and a sparse matrix $S$ via robust principal component analysis (RPCA) [37,38], where the low-rank matrix $L$ can be regard as a composition of the separated background patches.
$B_1 - B_l$, and the sparse matrix $S$ can be regard as a composition of the separated target patches $T_1 - T_l$, as shown in Figure 2.

![Figure 2. Local foreground-background images separation via RPCA.](image)

Compared with the conventional principal component analysis which can offer low-rank representation of a given data matrix corrupted by small i.i.d. Gaussian noise, the RPCA can work well with respect to grossly corrupted observations, and then the low-rank and sparse matrix decomposition can be achieved by many kinds of optimization techniques [39], in which the optimization problem can be formulated as following

$$\min_{L,S} (\text{rank}(L) + \lambda \|\text{vec}(S)\|_0) \quad \text{subject to} \quad G = L + S$$

(14)

where $\text{rank}(\cdot)$ is a function to solve the rank of matrix, $\text{vec}(\cdot)$ is vectorization for a matrix, $\|\cdot\|_0$ is the $\ell_0$-norm for vector which denotes the number of non-zero entries, and $\lambda$ is a weight parameter for tradeoff the low-rank matrix $L$ and sparse matrix $S$.

Conventionally, the optimization problem Equation (14) can be exactly solved via convex optimization, in which the rank function is relaxed to the nuclear norm $\|\cdot\|_*$ as its convex envelope, and the $\ell_0$-norm is relaxed to the $\ell_1$-norm as its convex envelope under a certain condition. Consequently, the optimization problem is converted to minimize a combination of the nuclear norm and the $\ell_1$-norm [40].

In this study, considering that each column of the data matrix $G$ is the vectorized image and the images are obtained from a static camera observing a fixed scene, the rank of the expected low-rank matrix $L$ can be approximately estimated. On the other hand, the pixel number of foreground target is specific in a short time, so the sparseness of the expected sparse matrix $S$ is also appreciable. Therefore, the optimization problem Equation (14) can be recast into a combination of the rank-$R$ matrix and $K$-sparse matrix approximation problem

$$\min_{L,S} \|\text{vec}(G - (L + S))\|_2 \quad \text{subject to} \quad \text{rank}(L) \leq R, \quad \|\text{vec}(S)\|_0 \leq K$$

(15)

where $\|\cdot\|_2$ is the $\ell_2$-norm for vector.

To solve the NP-hard problem, a pursuits algorithm dubbed SpaRCS for sparse and low-rank decomposition via compressive sensing is adopted [41]. Note that, in this study, the linear operator $A : \mathbb{R}^{d \times 1} \rightarrow \mathbb{R}^N$ is vectorization of a matrix, and its adjoining operator $A^* : \mathbb{R}^N \rightarrow \mathbb{R}^{d \times 1}$ denotes rearrangement from a vector to a matrix. The low-rank matrix $L$ and the sparse matrix $S$ are iteratively estimated by constructing two bases and attempting matrix recovery restricted to the bases. At each
iteration, five steps: supports identification, supports merging, least squares estimation, supports pruning and residue updating are respectively processed to update its estimates of \( L \) and \( S \).

On one hand, the low-rank matrix \( L \) is approximated by a set of up to \( R \) rank-1 matrices arranged from its singular vectors, and the set of entire singular vectors is defined as

\[
\mathcal{O} = \{ (u_i, v_i) : (U, \Sigma, V) = \text{svd} (\cdot) \}
\]

where \( u_i \) and \( v_i \) are respectively the \( i^{th} \) column vector of matrices \( U \) and \( V \), and \( \text{svd}(\cdot) \) is a function to singular value decomposition (SVD) of a given matrix. Further assume that a subset of singular vectors \( \Omega \subset \mathcal{O} \), and let \( F_{\Omega}(\cdot) \) denotes the projection operator onto the subspace spanned by \( \Omega \), then the estimation of low-rank matrix can be converted to a optimization problem about Frobenius norm of the projection matrix [42].

On the other hand, the sparse matrix \( S \) is approximated by the largest \( K \)-term support set of the given matrix, and the set of total practicable terms is defined as

\[
\mathcal{Q} = \{ (i, j) : i = 1, 2, \ldots; j = 1, 2, \ldots l \}
\]

where \( (i, j) \) denotes the position of support term. Further assume that a subset of support terms \( \Omega \subset \mathcal{Q} \), then let \( H_{\Omega} \in \mathbb{R}^{d \times l} \) denotes a 0-1 matrix with entity

\[
h_{ij} = \begin{cases} 1, & \text{if } (i, j) \in \Omega \\ 0, & \text{otherwise} \end{cases} \quad (i = 1, 2, \ldots; j = 1, 2, \ldots l)
\]

Subsequently, the estimation of sparse matrix can be converted to a optimization problem about Frobenius norm of dot product matrix [43].

3.3. Comprehensive Scheme and Algorithm Implementation

Based on the aforementioned discussion, a new approach to flying small target detection for anti-UAV based on a Gaussian mixture model in a compressive sensing domain is designed. The approach includes two operation stages: background model training and flying small target detecting, in which the flying small target detecting stage is composed by three process parts: candidate patches identification, local foreground–background image separation and threshold segmentation over reconstructed sparse images. The comprehensive scheme of the proposed approach is depicted in Figure 3.

![Figure 3. Comprehensive scheme for flying small target detection.](image)

Correspondingly, the synthetic integrated detection approach to flying small target in image sequence can be presented based on all above discussed modules, and outlined as Algorithm 1.
Algorithm 1 Approach to flying small target detection based on GMM_CS.

**Input:** Flying small target image sequence

**Output:** Detection results of each image in sequence

1. Initialize $N_{\text{train}}, N_{\text{detect}}, s, m, \xi, \alpha, \beta, R, K, \epsilon$
2. for $i = 1$ to $N_{\text{train}}$
do
    3. Break image into patches by Equation (11) and compressive sensing by Equation (3)
    4. Make GMM_CS for each patch by Equations (7)–(10)
    5. end for
6. for $i = 1$ to $N_{\text{detect}}$
do
    7. Break image into patches by Equation (11) and compressive sensing by Equation (3)
    8. Identify candidate patches by Equations (7), (12), (13)
    9. for $j = 1$ to $N_{\text{patch}}$
do
        10. Designate successive patches and compose a data matrix $G$
        11. repeat
            12. Set $k = 0, \hat{L}_0 = 0, \hat{S}_0 = 0, \hat{\Omega}_L = \{\}, \hat{\Omega}_S = \{\}, P_0 = G$
            13. Update indicator $k = k + 1$
            14. Identify supports: $\hat{\Omega}_L' = \arg \max_{\Omega \subseteq \hat{\Omega}_L} \{\| F_{\Omega} (P_{k-1}) \|_F : |\Omega| \leq 2R \}$
            15. Merging: $\hat{\Omega}_L = \hat{\Omega}_L \cup \hat{\Omega}_L'$
            16. Estimating: $\hat{L}_k = \arg \min_L \{\| G - L - \hat{S}_{k-1} \|_F : L \in \text{span}(\hat{\Omega}_L) \}$
            17. Pruning: $\hat{\Omega}_L = \arg \max_{\Omega \subseteq \hat{\Omega}_L} \{\| F_{\Omega} (\hat{L}_k) \|_F : |\Omega| \leq R \}$, $\hat{L}_k = F_{\hat{\Omega}_L} (\hat{L}_k)$
            18. Update residue: $P_k = G - \hat{L}_k - \hat{S}_{k-1}$
            19. Identify supports: $\hat{\Omega}_S' = \arg \max_{\Omega \subseteq \hat{\Omega}_S} \{\| H_{\Omega} \cdot \hat{P}_k \|_F : |\Omega| \leq 2K \}$
            20. Merging: $\hat{\Omega}_S = \hat{\Omega}_S \cup \hat{\Omega}_S'$
            21. Estimating: $\hat{S}_k = \arg \min_S \{\| G - \hat{L}_k - S \|_F : S = H_{\Omega} \cdot \hat{P}_k, \Omega \subseteq \hat{\Omega}_S \}$
            22. Pruning: $\hat{\Omega}_S = \arg \max_{\Omega \subseteq \hat{\Omega}_S} \{\| H_{\Omega} \cdot \hat{S}_k \|_F : |\Omega| \leq K \}$, $\hat{S}_k = H_{\hat{\Omega}_S} \cdot \hat{S}_k$
            23. Update residue: $P_k = G - \hat{L}_k - \hat{S}_k$
        24. until $\| G - \hat{\tilde{L}}_k - \hat{\tilde{S}}_k \|_F / \| G \|_F \leq \epsilon$
    9. end for
6. Detect target over reconstructed sparse images by threshold segmentation
end for

4. Experimental Results and Analysis

4.1. Experiment Setting and Evaluation Metrics

In order to validate the proposed detection method of flying small target, two experiments are conducted on visible and infrared image sequences of flying UAVs derived from two challenging datasets. The image sequence for the first experiment has a 189-frame visible image with resolution $640 \times 480$ pixels acquired by a CCD visible light camera, selected from the video named ‘video_37.avi’ in the UAV dataset which was collected by Rozantsev et al. [11]. The video with 14 s and 25 frames per seconds is the only one drone video captured from a stationary camera. Due to some discontinuous frames being without a ground-truth, we choose the labeled 189-frame visible image of this video for the experiment. The image sequence for the second experiment has a 337-frame infrared image...
with resolution 640 × 480 pixels acquired by a longwave FLIR A655sc camera, selected from the sequence named ‘quadrocopter2’ in the LTIR dataset which was used in the Visual Object Tracking (VOT) challenge [44]. The image sequence is captured from a stationary camera. Due to the moving foreground of flying target not being salient with no obvious changes, we choose 337 frames from the original 1010 frames by a three-frame interval for the experiment.

Additionally, the true positives (TP), false positives (FP) and false negatives (FN) of flying UAVs are counted, and then two most important evaluation metrics: the recall and precision are calculated to quantitatively evaluate the performance of the proposed flying small target detection method, and they are defined as follows:

\[
\text{Recall} = \frac{TP}{TP + FN} \tag{19}
\]

\[
\text{Precision} = \frac{TP}{TP + FP} \tag{20}
\]

In the evaluation, we regarded the overlap score between detected bounding boxes \(r_d\) and manually labeled ground-truth bounding boxes \(r_t\) that are greater than a certain threshold as correct detections and boxes that less than a certain threshold as misdetection, and the overlap score is defined as

\[
\text{Score} = \frac{|r_d \cap r_t|}{|r_d \cup r_t|} \tag{21}
\]

where \(\cap\) and \(\cup\) represent the intersection and union of two regions, respectively. \(|·|\) denotes the number of pixels in the region.

All the experiments are carried out on a general computer with 8-GB memory and 2.2-GHz Intel Core i5-5200U processor.

4.2. Performance Evaluation and Comparative Analysis

To test the effectiveness and practicality of the proposed method, its performance is evaluated by the aforementioned metrics and compared with three baseline methods. One is a well-studied background subtraction method based on an improved adaptive Gaussian mixture model (GMMv1) [45], another is a superior foreground and background separation method named Grassmannian robust adaptive subspace tracking algorithm (GRASTA) [46]. The third one is a state-of-the-art moving target detection method by detecting contiguous outliers in the low-rank representation (DECOLOR) [47].

The GMMv1 method is performed in a background subtraction library named BGSLibrary developed by Sobral [48], and its parameter setting in the library is adopted. The GRASTA and DECOLOR methods are implemented as open source and provided in this paper, and 20 frames visible or infrared image are introduced to train the initial subspace of GRASTA method, while the parameters of DECOLOR method employ the default values provided by the author.

The proposed method is implemented and executed on a Matlab R2012a software platform. In the following experiments, the number of Gaussian distribution is \(s = 3\). Correspondingly, the learning rate of weight is \(\alpha = 0.2\) and the learning rate of mean and variance is \(\beta = 0.2\) for each Gaussian distribution. In the procedure of CS measurement, the compressive ratio is \(m = 0.5n\). Moreover, the rank of the expected low-rank matrix \(R\) is set to 1, and the sparseness of the expected sparse matrix \(K\) is set to 0.1 times the number of pixels in one image patch. For identifying a image patch with a flying small targets, the threshold of \(\delta\) is set to 0.25\(m\) and 0.75\(m\) for the visible and infrared image sequences, respectively. In addition, 20 frames of visible or infrared images are introduced to train the proposed GMM_CS, and all images are broken into non-overlapping patches with a resolution of 80 × 60 and 64 × 48 pixels for the visible and infrared image sequences, respectively.

In the first experiment, a visible image sequences of a flying drone is tested to detect using three baseline methods and the proposed method. Due to the changing flight attitude, near or far distance and variable lighting conditions, the shape and size of the drone is in flux and is challenging for
detection. As an intuitively illustration, the flying small target detection results of four in-between frames are given as shown in Figure 4. The yellow number marked in the lower left quarter of image is the serial number of the image frame, and the target region of groundtruth, detected region by GMMv1, GRASTA, DECOLOR and the proposed method in this study are marked by a red bounding box. These results show that the proposed method can effectively detect the actual flying small target and generate less false alarm targets relative to the baseline methods.

Figure 4. Some experiment results of flying small target detection on visible image sequence obtained from different algorithms (a) groundtruth, (b) GMMv1, (c) GRASTA, (d) DECOLOR and (e) the proposed method.

In order to objectively quantitate the evaluation, the recall and precision are statistically calculated over the whole visual small target sequences. The thresholds of overlap score for target success detection affirmation are choose from 0.05 to 1 with step 0.05, and the result curves are plotted in Figure 5. It can be observed visually that the higher threshold of overlap score leads to lower recall and precision. Note that the proposed method outperforms the baseline methods in precision evaluation while they have approximately equal recall evaluation.

Figure 5. Curves of recall and precision with different threshold of overlap scores for flying small target detection on visible image sequence (a) recall and (b) precision.

In the second experiment, a longwave infrared image sequences of flying quadrocopter is tested to detect by three baseline methods and the proposed method. Due to the sequence coming from
a surveillance scenario with long imaging distance, the flying quadrocopter presents not only small size but also dim intensity even immersing in complex background. so it is rarely perfectly visible by the naked eye, not to mention automatic detection. For clear performance demonstration, the flying small target detection results in four in-between frames that are given as shown in Figure 6. The yellow number marked in the lower left quarter of the image is the serial number of the image frame, and the target region of groundtruth, detected region by GMMv1, GRASTA, DECOLOR and the proposed method in this study are marked by a red bounding box. We can find that the actual flying small target still can be detected in the challenging context in spite of being about some false alarm targets. Especially, the detection performance of the proposed method is not inferior to any baseline methods.

![Figure 6. Some experiment results of flying small target detection on infrared image sequence obtained from different algorithms (a) groundtruth, (b) GMMv1, (c) GRASTA, (d) DECOLOR and (e) the proposed method.](image-url)

Consequently, the recall and precision are statistically calculated over the whole infrared small target sequences. The thresholds of overlap score for target success detection affirmation are choose from 0.05 to 1 with step 0.05, and the resultant curves are plotted in Figure 7. It is easy to observed that the proposed method has better recall and precision evaluation when the thresholds of overlap score is less than 0.45, and has similar recall and precision evaluation when the thresholds of overlap score is more than 0.45.

![Figure 7. Curves of recall and precision with different threshold of overlap scores for flying small target detection on infrared image sequence (a) recall and (b) precision.](image-url)
5. Conclusions

In this paper, a new Gaussian mixture model in a compressive sensing domain is designed, and then a flying small target detection method for anti-UAV is proposed. The method not only makes use of the advantages of compressive sensing to reduce data quantity of background modeling, but also achieves flying small target detection on a handful of candidate patches by low-rank and sparse matrix decomposition. The results of experiments show that the proposed method can effectively detect flying small targets with less false alarms, and has outstanding recall and precision performance on both visible and infrared image sequences. In future studies, we will aim to focus on distributed deployment of multiple visual sensors for barrier coverage of surveillance, as well as multi-view stereo vision for flying small target detection by multi-sensor cooperation.

Author Contributions: C.W. proposed the idea of this paper and conducted the theoretical analysis and experimental work. T.W. proposed many useful suggestions in the process of theoretical analysis. E.S. and Z.L. took part in experiments. C.W. wrote the paper, and T.W. revised it.

Funding: This research was supported by the National Natural Science Foundation of China (Grant Nos. 61703287, 61503017 and 61571309), the Liaoning Provincial Natural Science Foundation of China (Grant No. 20180550664), the Scientific Research Program of Liaoning Provincial Education Department of China (Grant Nos. L201726 and L201702), the Science and Technology Plan Projects of Shenyang City of China (Grant No. 18-013-0-24) and the Doctoral Scientific Research Foundation of Shenyang Aerospace University (Grant No. 17YB16).

Acknowledgments: The authors would like to thank the reviewers and editors for their valuable comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, T.; Qin, R.; Chen, Y.; Snoussi, H.; Choi, C. A reinforcement learning approach for UAV target searching and tracking. Multimedia Tools Appl. 2019, 78, 4347–4364. [CrossRef]
2. Ganti, S.R.; Kim, Y. Implementation of detection and tracking mechanism for small UAS. In Proceedings of the 2016 International Conference on Unmanned Aircraft Systems (ICUAS), Arlington, VA, USA, 7–10 June 2016; pp. 1254–1260.
3. Mian, A.S. Realtime visual tracking of aircrafts. In Proceedings of the Digital Image Computing: Techniques and Applications (DICTA’08), Canberra, ACT, Australia, 1–3 December 2008; pp. 351–356.
4. Farlik, J.; Kratky, M.; Casar, J.; Stary, V. Radar cross section and detection of small unmanned aerial vehicles. In Proceedings of the 2016 17th International Conference on Mechatronics-Mechatronika (ME), Prague, Czech Republic, 7–9 December 2016; pp. 1–7.
5. Hoffmann, F.; Ritchie, M.; Fioranelli, E.; Charlish, A.; Griffiths, H. Micro-Doppler based detection and tracking of UAVs with multistatic radar. In Proceedings of the 2016 IEEE Radar Conference (RadarConf), Philadelphia, PA, USA, 2–6 May 2016; pp. 1–6.
6. Tapsall, T.B. Using crowd sourcing to combat potentially illegal or dangerous UAV operations. In Unmanned/Unattended Sensors and Sensor Networks XII; International Society for Optics and Photonics: Edinburgh, UK, 2016; Volume 9986, p. 998605.
7. Shi, X.; Yang, C.; Xie, W.; Chao, L.; Shi, Z.; Chen, J. Anti-Drone System with Multiple Surveillance Technologies: Architecture, Implementation, and Challenges. IEEE Commun. Mag. 2018, 56, 68–74. [CrossRef]
8. Sosnowski, T.; Bieszcadz, G.; Madura, H.; Kastek, M. Thermovision system for flying objects detection. In Proceedings of the 2018 Baltic URSI Symposium (URSI), Poznan, Poland, 15–17 May 2018; pp. 141–144.
9. Hwang, K.; Har, D. Identifying Small Drones Using Infrared Polarization Filters. TECHART J. Arts Imaging Sci. 2016, 3, 50–54. [CrossRef]
10. Li, J.; Ye, D.H.; Chung, T.; Kolsch, M.; Wachs, J.; Bouman, C. Multi-target detection and tracking from a single camera in Unmanned Aerial Vehicles (UAVs). In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Daejeon, South Korea, 9–14 October 2016; pp. 4992–4997.
11. Rozantsev, A.; Lepetit, V.; Fua, P. Flying objects detection from a single moving camera. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR2015), Boston, MA, USA, 7–12 June 2015; pp. 4128–4136.

12. Rozantsev, A.; Lepetit, V.; Fua, P. Detecting Flying Objects using a Single Moving Camera. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2017, 39, 879–892. [CrossRef] [PubMed]

13. Zhang, Z.; Cao, Y.; Ding, M.; Zhuang, L.; Wang, Z. Spatial and Temporal Context Information Fusion Based Flying Objects Detection for Autonomous Sense and Avoid. In Proceedings of the 2018 International Conference on Unmanned Aircraft Systems (ICUAS), Dallas, TX, USA, 12–15 June 2018; pp. 569–578.

14. Du, L.; Gao, C.; Feng, Q.; Wang, C.; Liu, J. Small UAV Detection in Videos from a Single Moving Camera. In Proceedings of the CCF Chinese Conference on Computer Vision, Tianjin, China, 11–14 October 2017; Springer: Singapore, 2017; pp. 187–197.

15. Wu, P.F.; Xiao, F.; Sha, C.; Huang, H.P.; Wang, R.C.; Xiong, N.X. Node Scheduling Strategies for Achieving Full-View Area Coverage in Camera Sensor Networks. *Sensors* 2017, 17, 1303. [CrossRef]

16. Han, R.; Yang, W.; Zhang, L. Achieving Crossed Strong Barrier Coverage in Wireless Sensor Network. *Sensors* 2018, 18, 534. [CrossRef] [PubMed]

17. Rossi, P.S.; Ciuonzo, D.; Kansanen, K.; Ekman, T. On energy detection for MIMO decision fusion in wireless sensor networks over NLOS fading. *IEEE Commun. Lett.* 2015, 19, 303–306. [CrossRef]

18. Rossi, P.S.; Ciuonzo, D.; Kansanen, K.; Ekman, T. Performance analysis of energy detection for MIMO decision fusion in wireless sensor networks over arbitrary fading channels. *IEEE Trans. Wirel. Commun.* 2016, 15, 7794–7806. [CrossRef]

19. Zhuang, X.; Huang, J.; Potamianos, G.; Hasegawa-Johnson, M. Acoustic fall detection using Gaussian mixture models and GMM supervectors. In Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing, Taipei, Taiwan, 19–24 April 2009; pp. 69–72.

20. Wang, T.; Qiao, M.; Chen, J.; Wang, C.; Zhang, W.; Snoussi, H. Abnormal global and local event detection in compressive sensing domain. *IEEE Commun. Lett.* 2015, 19, 871–874. [CrossRef] [PubMed]

21. Stein, D.W. Detection of random signals in Gaussian mixture noise. *IEEE Trans. Inf. Theory* 1995, 41, 1788–1801. [CrossRef]

22. Wang, T.; Qiao, M.; Lin, Z.; Li, C.; Gao, F. Infrared small target detection in compressive domain. *Electron. Lett.* 2014, 50, 510–512. [CrossRef] [PubMed]

23. Wang, C.; Qin, S. Background Modeling of Infrared Image in Dynamic Scene with Gaussian Mixture Model in Compressed Sensing Domain. *Acta Autom. Sin.* 2018, 44, 1212–1226.

24. Wang, T.; Qiao, M.; Lin, Z.; Li, C.; Snoussi, H.; Liu, Z.; Choi, C. Generative Neural Networks for Anomaly Detection in Crowded Scenes. *IEEE Trans. Inf. Forensics Secur.* 2019, 14, 1390–1399. [CrossRef]
35. Gan, L. Block Compressed Sensing of Natural Images. In Proceedings of the International Conference on Digital Signal Processing, Cardiff, UK, 1–4 July 2007; pp. 403–406.
36. Cho, T.S.; Butman, M.; Avidan, S.; Freeman, W.T. The patch transform and its applications to image editing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR2008), Anchorage, AK, USA, 23–28 June 2008; pp. 1–8.
37. Candès, E.J.; Li, X.; Ma, Y.; Wright, J. Robust principal component analysis? J. ACM 2011, 58, 1–39. [CrossRef]
38. Chen, C.; Li, S.; Qin, H.; Hao, A. Robust salient motion detection in non-stationary videos via novel integrated strategies of spatio-temporal coherency cues and low-rank analysis. Pattern Recognit. 2016, 52, 410–432. [CrossRef]
39. Wright, J.; Ganesh, A.; Rao, S.; Ma, Y. Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Matrices via Convex Optimization. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 7–10 December 2009; pp. 2080–2088.
40. Wang, C.; Qin, S. Approach for moving small target detection in infrared image sequence based on reinforcement learning. J. Electron. Imaging 2016, 25, 053032. [CrossRef]
41. Waters, A.E.; Sankaranarayanan, A.C.; Baraniuk, R. SpaRCS: Recovering low-rank and sparse matrices from compressive measurements. In Proceedings of the Advances in Neural Information Processing Systems, Granada, Spain, 12–14 December 2011; pp. 1089–1097.
42. Lee, K.; Bresler, Y. ADMiRA: Atomic decomposition for minimum rank approximation. IEEE Trans. Inf. Theory 2010, 56, 4402–4416. [CrossRef]
43. Needell, D.; Tropp, J.A. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. Appl. Comput. Harmon. Anal. 2009, 26, 301–321. [CrossRef]
44. Berg, A.; Ahlberg, J.; Felsberg, M. A thermal Object Tracking benchmark. In Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance, Karlsruhe, Germany, 25–28 August 2015; pp. 1–6.
45. Kaewtrakulpong, P.; Bowden, R. An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection. In Video Based Surveillance Systems; Remagnino, P., Jones, G.A., Paragios, N., Regazzoni, C.S., Eds.; Springer: Boston, MA, USA, 2002; pp. 135–144.
46. He, J.; Balzano, L.; Szlam, A. Incremental gradient on the Grassmannian for online foreground and background separation in subsampled video. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR2012), Providence, RI, USA, 16–21 June 2012; pp. 1568–1575.
47. Zhou, X.; Yang, C.; Yu, W. Moving Object Detection by Detecting Contiguous Outliers in the Low-Rank Representation. IEEE Trans. Pattern Anal. Mach. Intell. 2013, 35, 597–610. [CrossRef]
48. Sobral, A. BGSLibrary: An opencv c++ background subtraction library. In Proceedings of the IX Workshop De Visão Computacional, Botafogo, Rio de Janeiro, Brazil, 3–5 June 2013; pp. 1–6.

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).